Evaluation der Maturitätsreform 1995 (EVAMAR)

Schlussbericht zur Phase II

Franz Eberle, Karin Gehrer, Beat Jaggi, Johannes Kottonau, Maren Oepke, Michael Pflüger;

Institut für Gymnasial- und Berufspädagogik, UZH
(Kernteam der Evaluation)

Christina Huber, Vera Husfeldt, Lukas Lehmann,
Carsten Quesel

Fachhochschule Nordwestschweiz,
Pädagogische Hochschule
(Autoren des Kapitels VI)

Im Auftrag der
Schweizerischen Konferenz der kantonalen Erziehungsdirektoren EDK
und des Staatssekretariats für Bildung und Forschung SBF

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Staatssekretariat für Bildung und Forschung SBF

Das Staatssekretariat für Bildung und Forschung veröffentlicht in seiner „Schriftenreihe SBF“ konzeptionelle Arbeiten, Forschungsergebnisse und Berichte zu aktuellen Themen in den Bereichen Bildung und Forschung, die damit einem breiteren Publikum zugänglich gemacht und zur Diskussion gestellt werden sollen.

Die präsentierten Analysen geben nicht notwendigerweise die Meinung des Staatssekretariats für Bildung und Forschung wieder.

© 2008 Staatssekretariat für Bildung und Forschung SBF
ISSN: 1662 - 2634

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Staatssekretariat für Bildung und Forschung SBF
Bereich Bildung

Hallwylstrasse 4
CH - 3003 Bern

T +41 31 322 96 69
F +41 31 322 78 54

info@sbf.admin.ch
www.sbf.admin.ch

Evaluation der Maturitätsreform 1995 (EVAMAR)

Schlussbericht zur Phase II

Franz Eberle, Karin Gehrer, Beat Jaggi, Johannes Kottonau, Maren Oepke, Michael Pflüger;

Institut für Gymnasial- und Berufspädagogik, UZH
(Kernteam der Evaluation)

Christina Huber, Vera Husfeldt, Lukas Lehmann,
Carsten Quesel

Fachhochschule Nordwestschweiz,
Pädagogische Hochschule
(Autoren des Kapitels VI)

Im Auftrag der
Schweizerischen Konferenz der kantonalen Erziehungsdirektoren EDK
und des Staatssekretariats für Bildung und Forschung SBF

Inhalt

Vorwort des Projektleiters 13

Das Wichtigste in Kürze 15

I Einleitung zum Gesamtprojekt 21

1 Reform der Maturitätsbildung 21

2 EVAMAR I 23

3 Weitere Evaluationsstudien 24

4 Teilrevision des MAR 95 25

5 Auftrag und Fragestellung von EVAMAR II 26

5.1 Das umfassende Untersuchungskonzept 26

5.2 Die für EVAMAR II ausgewählten bzw. in Auftrag gegebenen Untersuchungen 29

5.3 Grenzen der Studie EVAMAR II 31

5.4 Projektorganisation und Dank 33

II Projektteil A: Elemente der Studierfähigkeit aus Sicht der Hochschulen 35

1 Datenbasis der Teilprojekte TP A1, TP A3 und TP A4 36

2 Teilprojekte A1 und A3: Inhaltsanalyse von Studienmaterialien und Prüfungen 37

2.1 Methode und Instrumente TP A1 und TP A3 37

2.1.1 Die Methode des Feincodierens 37

2.1.2 Die Methode des grobmaschigen Codierens 40

2.1.3 Zur messtechnischen Qualität 41

2.2 Datenrücklauf in TP A1 und TP A3 43

2.3 Auswertungen in TP A1 und TP A3 43

2.4 Auswertungsergebnisse TP A1 und TP A3 und Interpretation 44

3 Teilprojekt A4: Dozierendenbefragung 45

3.1 Der Fragebogen 45

3.2 Datenbasis von TP A4 und Ablauf der Befragung 46

3.3 Ergebnisse der Befragung in TP A4 46

3.3.1 Bedeutung der gymnasialen Fachbereiche für die Lehrveranstaltungen (Frage 1) 46

3.3.2 Konkrete Wissensinhalte aus den für die Lehrveranstaltung bedeutsamen gymnasialen Fachbereichen (Frage 2) 50

3.3.3 Aktuell festgestellte Wissens- und Könnenslücken (Frage 3) 53

3.3.4 Wichtigkeit überfachlicher Kompetenzen für die Lehrveranstaltung (Frage 4) 55

3.3.5 Festgestellte überfachliche Wissens- und Könnensdefizite (Frage 5) 57

3.3.6 Wünschbares zusätzliches Eingangswissen und -können 57

3.3.7 Schlussfolgerungen 60

4 Teilprojekt A2: Analyse von Eingangsprüfungen **61**
4.1 Übersicht 61
4.2 Der Medizineignungstest 61
4.3 Die Aufnahmeprüfung an der HSG 62
4.4 Schlussfolgerungen 62

III Projektteil B: Modellierung hochschulrelevanten Wissens und Könnens in einzelnen Fächern 63

1 Allgemeines **63**

2 Der Kompetenzraster in Erstsprache **64**
2.1 Grundlagen 64
2.2 Bestehende Kompetenzraster und -modelle für Erstsprache 65
2.2.1 Zürcher Textanalyseraster 65
2.2.1.1 Kurzbeschreibung 65
2.2.2 DESI 66
2.2.2.1 Kurzbeschreibung 66
2.2.2.2 Einschätzung des Kompetenzmodells von DESI für die Verwendung in EVAMAR II 68
2.2.3 PISA 68
2.2.3.1 Kurzbeschreibung 68
2.2.3.2 Einschätzung des PISA-Modells für die Verwendung in EVAMAR II 71
2.2.4 Der Gemeinsame Europäische Referenzrahmen (GER) 71
2.2.4.1 Kurzbeschreibung 71
2.2.4.2 Einschätzung des GER für die Verwendung in EVAMAR II 75
2.2.5 HarmoS 76
2.2.5.1 Kurzbeschreibung 76
2.2.5.2 Einschätzung des HarmoS-Modells für die Verwendung in EVAMAR II 78
2.2.6 Lernstandserhebung in Nordrhein-Westfalen (NRW) 78
2.2.6.2 Einschätzung der Lernstandserhebungen NRW für die Verwendung in EVAMAR II 79
2.3 Das Endmodell „Kompetenzraster Erstsprache“ für EVAMAR II 80
2.3.1 Begründung der Wahl des GER als Grundlage 80
2.3.2 Beschreibung des EVAMAR-Kompetenzrasters für Erstsprache 81
2.4 Zusammenfassung 84

3 Der Kompetenzraster in Mathematik **86**
3.1 Grundlagen 86
3.2 Bestehende Kompetenzmodelle und Aufgabenraster für Mathematik 87
3.2.1 Das Kompetenzmodell des NCTM (National Council of Teachers of Mathematics) 87
3.2.2 Das Kompetenzmodell der KMK in Deutschland 87
3.2.3 Das Kompetenzmodell für Mathematik bei HarmoS 88
3.2.4 Das Kompetenzmodell für Mathematik bei PISA 90
3.2.5 Modelle bei TIMSS 90
3.3 Das Endmodell „Kompetenzraster Mathematik“ für EVAMAR II 92

4 Der Kompetenzraster in Biologie **95**
4.1 Grundlagen 95
4.2 Bestehende Kompetenzraster und -modelle für Biologie 95
4.2.1 Die Regensburger Verbundklassifikation (RVK) 95
4.2.2 Vergleich der Kapiteleinteilung deutschsprachiger Standardlehrbücher auf der gymnasialen Stufe 96
4.2.3 Vergleich der Kapiteleinteilung von Standardlehrbüchern für Anfänger des Biologiestudiums 97

4.2.4 Die EPA Biologie und Bildungsstandards in Deutschland 100
4.2.4.1 EPA Biologie 100
4.2.4.2 Bildungsstandards KMK Biologie 102
4.2.5 Zwischenergebnis für die inhaltliche Ebene: Zusammenstellung 104
4.2.6 Scientific Literacy 108
4.2.7 Bildungsstandards KMK 108
4.2.8 EPA Biologie 109
4.3 Beschreibung des EVAMAR-Kompetenzrasters für Biologie 110

IV Projektteil C: Entwicklung und Durchführung von Tests 113

1 Allgemeines und Übersicht **113**
1.1 Zielsetzungen und Auswahl der Erhebungsgegenstände 113
1.2 Präzisierungen zur Bestimmung der Testbereiche 113
1.2.1 Zur Prognosevalidität der Tests im Hinblick auf die Studierfähigkeit 113
1.2.2 Zur Wahl von Biologie als naturwissenschaftliches Fach 118

2 Beschreibung der Erhebungsinstrumente **119**
2.1 Vorgehen zur Konstruktion der Tests: Grundsätzliches für alle Tests 119
2.2 Beschreibung der qualitativen Expertenevaluation und der Pretests (inkl. Verfahren und Überarbeitung der Items) 121
2.3 Die Übersetzungen in die zwei anderen Landessprachen: Grundsätzliches für alle drei Fachgebiete 122
2.4 Vorgehen zur Auswahl der in die Hauptauswertungen einbezogenen Items 122
2.5 Der Test in Erstsprache 123
2.5.1 Ergänzendes zur Testentstehung 123
2.5.1.1 Besonderes und Beispiel zur Aufgabenkonstruktion 123
2.5.1.2 Besonderes bei der Übersetzung 127
2.5.2 Beschreibung der Testhefte 128
2.5.3 Charakterisierung der in die Hauptauswertungen einbezogenen Aufgaben und Verankerung im Kompetenzraster 129
2.6 Der Test in Mathematik 129
2.6.1 Ergänzendes zur Testentstehung 129
2.6.1.1 Besonderes und Beispiel zur Aufgabenkonstruktion 129
2.6.1.2 Besonderes bei der Übersetzung 131
2.6.2 Beschreibung der Testhefte 131
2.6.3 Charakterisierung der in die Hauptauswertungen einbezogenen Aufgaben und Verankerung im Kompetenzraster 132
2.7 Der Test in Biologie 132
2.7.1 Ergänzendes zur Testentstehung 132
2.7.1.1 Besonderes und Beispiel zur Aufgabenkonstruktion 132
2.7.1.2 Besonderes bei der Übersetzung 136
2.7.2 Beschreibung der Testhefte 136
2.7.3 Charakterisierung der in die Hauptauswertungen einbezogenen Aufgaben und Verankerung im Kompetenzraster 137
2.8 Der überfachliche Fähigkeitstest (UFT) 137
2.8.1 Beschreibung der Grundlagen 137
2.8.2 Beschreibung der Testhefte 138
2.8.3 Charakterisierung der in die Hauptauswertungen einbezogenen Aufgaben 138
2.9 Der Zusatzfragebogen 138
2.9.1 Begründung der Zusatzerhebung und Inhalte 138
2.9.2 Beschreibung des Fragebogens 139
2.9.3 Güte des Instruments 140

3 Bestimmung der Stichprobe **140**
3.1 Bestimmung der Grundgesamtheit 140
3.2 Das Stichprobenverfahren 141
3.3 Rücklaufquote 142

4 Ablauf der Datenerhebung, Dateneingabe, Intercoder-Reliabilität **143**
4.1 Ablauf der Datenerhebung 143
4.2 Codierungsverfahren bei Erstsprache 143

5 Nationale Ergebnisse **143**
5.1 Allgemeines 143
5.2 Ergebnisse in Erstsprache 145
5.2.1 Beschreibung der Gesamtergebnisse 145
5.2.2 Beschreibung der Ergebnisse ausgewählter Items 146
5.3 Ergebnisse in Mathematik 154
5.3.1 Beschreibung der Gesamtergebnisse 154
5.3.2 Beschreibung der Ergebnisse ausgewählter Items 156
5.4 Ergebnisse in Biologie 161
5.4.1 Beschreibung der Gesamtergebnisse 161
5.4.2 Beschreibung der Ergebnisse ausgewählter Items 162
5.5 Zum Vergleich: Erzielte Maturanoten 170
5.6 Ergebnisse „Überfachlicher Fähigkeitstest“ (UFT) 171
5.7 Ergebnisse Zusatzfragebogen 172
5.7.1 Einschätzung der Tests 172
5.7.2 Wahrnehmung überfachlicher Kompetenzen 173
5.7.3 Fachinteressen und Studienpropädeutik (nur Deutschschweiz) 174
5.8 Interpretation der Ergebnisse im Hinblick auf die Studierfähigkeit 175

6 Vergleichende Auswertungen **177**
6.1 Grundsätzliches 177
6.2 Vergleich zwischen den Straten (Gruppen von Gymnasien) 178
6.2.1 Vergleichende Testergebnisse für die Straten 178
6.2.1.1 Erstsprache 178
6.2.1.2 Mathematik 180
6.2.1.3 Biologie 182
6.2.2 Vergleich der Maturanoten zwischen den Straten 185
6.2.3 Vergleich der überfachlichen Fähigkeiten zwischen den Straten 187
6.2.4 Vergleich ausgewählter Ergebnisse aus der Zusatzbefragung 188
6.3 Vergleich nach Schwerpunktfächern 190
6.3.1 Vergleich der Fach-Testergebnisse nach Schwerpunktfächern 190
6.3.1.1 Erstsprache 190
6.3.1.2 Mathematik 193
6.3.1.3 Biologie 195
6.3.2 Vergleich der Maturanoten nach Schwerpunktfächern 199
6.3.3 Vergleich der überfachlichen Fähigkeiten nach Schwerpunktfächern 201
6.3.4 Vergleich ausgewählter Ergebnisse aus der Zusatzbefragung nach Schwerpunktfächern 203
6.4 Vergleich zwischen den Geschlechtern 205
6.4.1 Geschlechtervergleich in Erstsprache 205
6.4.2 Geschlechtervergleich in Mathematik 205
6.4.3 Geschlechtervergleich in Biologie 206
6.4.4 Geschlechtervergleich bei den überfachlichen Fähigkeiten 207
6.4.5 Vergleich ausgewählter Ergebnisse aus der Zusatzbefragung 208
6.5 Vergleich zwischen unterschiedlichen Maturitätsquoten 208

6.6 Vergleiche nach Strukturelement Lang- oder Kurzzeitgymnasium 211
6.7 Weitere Vergleiche 213
6.7.1 Anstrengungsbereitschaft und Testergebnisse 213
6.7.2 Vergleich von mündlichen und schriftlichen Noten 215

7 Ausgewählte korrelative Zusammenhänge **216**
7.1 Zusammenhänge zwischen den Tests 216
7.2 Zusammenhänge der Tests mit den Maturanoten 217
7.3 Zusammenhänge der Maturanoten untereinander 218

8 Zusammenfassung **219**
8.1 Grundlagen der Test- und Befragungsergebnisse 219
8.2 Gesamtergebnisse der Tests 219
8.3 Ergebnisse der Tests unter strukturellen Gesichtspunkten 220
8.4 Notengebung 221
8.5 Eigene Wahrnehmung überfachlicher Kompetenzen 222

V Projektteil D1: Analyse schriftlicher Maturaprüfungen 223

1 Allgemeines und Übersicht **223**
1.1 Zielsetzungen und Auswahl der Fächer 223
1.2 Allgemeines zu den Analyse-Instrumenten 223

2 Beschreibung der Stichprobe **227**

3 Ergebnisse **227**
3.1 Erstsprache 227
3.1.1 Das Analyse-Instrumentarium 227
3.1.2 Die Prüfungsinhalte in Erstsprache 228
3.1.2.1 Die Aufgabenform 228
3.1.2.2 Der vorgelegte Primärtext 230
3.1.2.3 Zeitbezug 233
3.1.2.4 Autorinnen und Autoren 234
3.1.2.5 Themen 238
3.1.2.6 Länge der vorgelegten Texte 238
3.1.2.7 Wahlmöglichkeiten 240
3.1.2.8 Unterstützende Hinweise 240
3.1.2.9 Hilfsmittel 241
3.1.2.10 Dauer der Prüfung 242
3.1.3 Kognitive Anforderungen 242
3.1.4 Vergleich mit den Aufgaben in Teilprojekt C 245
3.1.5 Vergleich mit den Ergebnissen aus Teilprojekt A4 246
3.1.6 Gesamteinschätzung der Schweizer Maturaprüfungen in Erstsprache 248
3.2 Mathematik 249
3.2.1 Das Analyse-Instrumentarium 249
3.2.2 Die Prüfungsinhalte in Mathematik 250
3.2.2.1 Vorgehen 250
3.2.2.2 Ergebnisse 251
3.2.3 Kognitive Anforderungen 252
3.2.4 Vergleich mit den Aufgaben in Teilprojekt C 258
3.2.5 Vergleich mit den Ergebnissen aus Teilprojekt A4 258
3.2.6 Gesamteinschätzung der Schweizer Maturaprüfungen in Mathematik 261
3.3 Biologie 261
3.3.1 Das Analyseinstrumentarium 261
3.3.2 Prüfungsinhalte in Biologie 262

3.3.3 Kognitive Anforderungen 264
3.3.4 Vergleich mit den Aufgaben in Teilprojekt C 271
3.3.5 Vergleich mit den Ergebnissen aus Teilprojekt A4 273
3.3.6 Gesamteinschätzung der Schweizer Maturaprüfungen in Biologie 274

VI Teilprojekt D2: Die Qualität von Maturaarbeiten in der Schweiz 277

1 Ausgangslage und theoretischer Hintergrund **278**
1.1 Die Maturaarbeit als Teil der Maturitätsreform 1995 278
1.1.1 Einführung der Maturaarbeit 278
1.1.2 Idee der Maturaarbeit als interdisziplinäre Projektarbeit 279
1.1.3 Die Maturaarbeit im Prüfungskontext 279
1.2 Die Maturaarbeit im Kontext gymnasialer Bildung 280
1.2.1 Studierfähigkeit und Hochschulreife 281
1.2.2 Wissenschaftspropädeutik 282
1.2.3 Überfachliche Kompetenzen 284
1.2.4 Zwischenbilanz 285
1.3 Forschungsstand zur Thematik der Maturaarbeiten 285
1.3.1 Typen von Maturaarbeiten 286
1.3.2 Themenwahl 286
1.3.3 Durchführung 287
1.3.4 Wissen, Fähigkeiten und Fertigkeiten der Schülerinnen und Schüler 287
1.3.5 Betreuung 288
1.3.6 Zwischenbilanz 288

2 Methodische Grundlagen und Durchführung der Untersuchung **289**
2.1 Wirkungsmodell 289
2.2 Beurteilung der Maturaarbeiten 290
2.2.1 Zuschnitt der Untersuchung 290
2.2.2 Bezugsnormen zur Beurteilung der Qualität von Maturaarbeiten 290
2.2.3 Entwicklung des Analyseinstrumentes 293
2.2.4 Ratingverfahren 293
2.2.5 Überprüfung von Plagiaten 294
2.3 Vergleichende Darstellung der Rahmenbedingungen 295
2.4 Befragung der Schülerinnen und Schüler 295
2.5 Überprüfung des postulierten Wirkungsgefüges 296
2.5.1 Qualität der Maturaarbeit 296
2.5.2 Institutionelle Ebene 297
2.5.3 Betreuungsebene 298
2.5.4 Individuelle Ebene 299
2.6 Untersuchungspopulation und Stichprobenziehung 301
2.6.1 Untersuchungspopulation 301
2.6.2 Stichprobenziehung 301
2.6.3 Datenerhebung 303

3 Institutionelle Rahmenbedingungen **304**
3.1 Übergeordnete Vorgaben 304
3.1.1 Zielsetzungen auf nationaler Ebene 304
3.1.2 Zielsetzungen auf kantonaler Ebene 305
3.2 Vorgaben auf der Ebene der Einzelschule 307
3.2.1 Zuständigkeiten und Richtlinien 307
3.2.2 Gymnasiale Zielbestimmungen 308
3.2.3 Vorbereitung und Themenwahl 309
3.2.4 Formale Vorgaben 311

3.2.5 Arbeitsprozess und Zeitpläne 312
3.2.6 Betreuung 313
3.2.7 Beurteilung und Bewertung 314
3.3 Zwischenbilanz 315

4 Die Maturaarbeit im Urteil der Schülerinnen und Schüler **316**
4.1 Fächerwahl 317
4.2 Fachbezug der Maturaarbeit und Themenwahl 317
4.3 Themenwahl bei der Maturaarbeit 319
4.4 Bewertung der Maturaarbeit 321
4.5 Betreuung 322
4.6 Institutionelle Rahmenbedingungen und individuelle Voraussetzungen 328
4.7 Einschätzungen zu Aufwand und Nutzen der Maturaarbeit 331
4.8 Zwischenbilanz 336

5 Die wissenschaftspropädeutische Qualität der Maturaarbeiten **336**
5.1 Erst- und Zweitratings 337
5.2 Angaben zur Stichprobe 338
5.3 Skalenbildung 339
5.4 Inhaltliche Qualität 342
5.5 Sprachliche Qualität 343
5.6 Formale Qualität 344
5.7 Globalurteil 346
5.8 Individuelle und institutionelle Qualitätsaspekte 346
5.9 Zwischenbilanz 350

6 Fazit **350**
Literatur- und Quellenverzeichnis zu Kapitel VI **353**
Tabellen und Abbildungen zu Kapitel VI **360**
Anhang zu Kapitel VI **362**
Anhang VI.A: Übersicht zu den Hypothesentests **362**
Anhang VI.B: Raster Rahmenbedingungen **364**
Anhang VI:C: Gewichtungsfaktoren **365**

VII Zusammenfassung 367

1 Fragestellungen **367**

2 Ergebnisse I: Voruntersuchung zur Konkretisierung des zu Messenden **368**

3 Kompetenzmodellierung – TP B **372**

4 Ergebnisse II: Leistungsmessung – TP C **372**

5 Ergebnisse III: Analyse der Maturaprüfungen **379**

6 Ergebnisse IV: Analyse der Maturaarbeiten **380**

7 Grenzen der Studie EVAMAR II **382**

VIII Schlussfolgerungen und Empfehlungen des Projektleiters 383

1 Zufriedenstellendes Gesamtergebnis **383**

2 Massnahmen zur Sicherung der allgemeinen Studierfähigkeit **384**

3 Beibehaltung der breiten Fächerung und der allgemeinen Zutrittsberechtigung an Hochschulen **387**

4 Ausgleich der Defizite zu Beginn eines Studiums **388**

5 Massnahmen zur Förderung des selbstständigen Lernens ohne Abbau der Fachlichkeit **388**

6 Massnahmen zur Verbesserung der Erstsprachkompetenzen **389**

7 Vermehrte Standardorientierung bei den Unterrichtsprogrammen **389**

8 Vermehrte Standardorientierung bei der Leistungsbewertung **389**

9 Beibehaltung der vierjährigen Ausbildungsdauer am Schulort Gymnasium **390**

10 Sicherung der allgemeinen Studierfähigkeit bei allen Schwerpunkten **391**

11 Annäherung der Maturitätsquoten **391**

12 Weitere Anstrengungen bei der Gender-Problematik **392**

13 Weiterentwicklung der Maturaarbeit **392**

14 Weitere Untersuchungen **392**

Literaturverzeichnis (ohne Kapitel VI) 393

Abbildungs- und Tabellenverzeichnis 405

Abkürzungsverzeichnis 412

Vorwort des Projektleiters

Im Sommer 2001 beschlossen der Bund und die EDK eine gesamtschweizerische Evaluation (EVAMAR) der durch das MAR 95 eingeleiteten Reform der Maturitätsbildung. In einer ersten Phase (EVAMAR I) wurden mittels Befragungen die Passung von Wahlfachangebot und Interessen der Schülerinnen und Schüler sowie die subjektiv wahrgenommene Qualität der Vorbereitung auf ein Hochschulstudium, die Umsetzung der fächerübergreifenden pädagogischen Ziele sowie die Bewältigung der Reformen durch die Schulen untersucht. Die Ergebnisse fielen zufriedenstellend aus. Die 2. Phase (EVAMAR II) begann Ende 2005 und wurde im Herbst 2008 abgeschlossen. In EVAMAR II wurden weitere Aspekte der Maturitätsbildung untersucht. Das Schwergewicht lag auf der objektivierten Erfassung des Ausbildungsstandes der Schülerinnen und Schüler am Ende des Gymnasiums im Hinblick auf die Anforderungen eines Hochschulstudiums. Auch wenn es kaum möglich ist, mit Genauigkeit die wirkliche Studierfähigkeit in ihrer ganzen Breite für alle Maturandinnen und Maturanden zu messen und zu prognostizieren, erlauben die vorliegenden Resultate Einblicke in ausgewählte Bereiche der Ergebnisse der Maturitätsbildung an Schweizer Gymnasien.

Ich danke an erster Stelle all den Maturandinnen und Maturanden, die bereit waren, die Tests zu absolvieren und die Fragebögen auszufüllen oder die uns ihre Maturaarbeiten für das Rating zur Verfügung gestellt haben, und ich danke insbesondere auch den zahlreichen Lehrpersonen, Mitgliedern von Schulleitungen und administrativen Mitarbeitenden, die uns bei der Erhebung der Daten aktiv unterstützt haben.

Ein ebenso grosser Dank gebührt den vielen am Projekt beteiligten Personen, die am Schluss des ersten Kapitels aufgeführt sind. Ohne ihre grosse Arbeit wäre die Durchführung eines so komplexen Projekts nicht möglich gewesen.

Der vorliegende Bericht wurden von verschiedenen Autorinnen und Autoren verfasst, die auch die entsprechenden Untersuchungen und Analysen durchgeführt haben. Deshalb finden sich zwischen einzelnen Kapiteln und Abschnitten stilistische und fachbezogen auch strukturelle Unterschiede in der Darstellung.

Das Kapitel VI wurde gänzlich von den Autoren und Autorinnen der Fachhochschule Nordwestschweiz, Pädagogische Hochschule, geschrieben. Sie haben dieses Teilprojekt konzipiert und durchgeführt.

Zürich, 31. Oktober 2008 Franz Eberle

Das Wichtigste in Kürze

Im Sommer 2001 beschlossen der Bund und die Konferenz der kantonalen Erziehungsdirektoren (EDK) eine gesamtschweizerische Evaluation (EVAMAR) der durch das MAR 95 (1995) eingeleiteten Reform der Maturitätsbildung. In der ersten Phase (EVAMAR I) wurden im Wesentlichen die folgenden drei Themen bearbeitet: (1) die Passung von Wahlfachangebot und Interessen der Schülerinnen und Schüler sowie die Auswirkungen auf den Ausbildungserfolg, insbesondere die subjektiv wahrgenommene Qualität der Vorbereitung auf ein Hochschulstudium; (2) die Umsetzung der fächerübergreifenden pädagogischen Ziele; (3) die Bewältigung der Reform durch die Schulen. Hauptinstrumente waren Befragungen, es erfolgten keine Leistungsmessungen. Die Ergebnisse von EVAMAR I liegen seit Ende 2004 vor. Sie bewerteten die Reform überwiegend als zufriedenstellend. Im Sommer 2005 beschlossen Bund und EDK den Beginn der zweiten Evaluationsphase. Im Projekt EVAMAR II sollte das Schwergewicht auf die objektivierte Erfassung des Ausbildungsstandes der Schülerinnen und Schüler am Ende des Gymnasiums gelegt werden.

Fragestellungen

Eine evaluative Erfassung des Ausbildungsstandes muss sich sinnvollerweise an den Zielen der Ausbildung orientieren. Die Hauptziele der gymnasialen Bildung sind gemäss Art. 5 des MAR 95 (1995) die Erlangung „jener persönlichen Reife“, die erstens „Voraussetzung für ein Hochschulstudium ist“ (Hochschulreife oder Studierfähigkeit), und die zweitens „auf anspruchsvolle Aufgaben in der Gesellschaft vorbereitet“ (vertiefte Gesellschaftsreife durch breite Allgemeinbildung). Diese beiden Ziele haben teils gleichgerichtete, teils aber auch eigenständige und mitunter breite curriculare Auswirkungen. Eine vollständige Operationalisierung dieser Ziele in konkret zu erwerbende Kompetenzen der Gymnasiastinnen und Gymnasiasten in allen fachlichen und überfachlichen Lernbereichen und eine entsprechende flächendeckende Messung des Ausbildungsstandes, wie sie in einem ersten Konzept des Auftragnehmers vorgeschlagen wurde, wäre ausserordentlich aufwändig. Deshalb sollte die Konkretisierung der Ziele der Maturitätsbildung nur für ausgewählte Elemente des Hauptzieles der Studierfähigkeit erfolgen. Das zweite Hauptziel (vertiefte Gesellschaftsreife durch breite Allgemeinbildung) wurde damit nur teilweise (im Überschneidungsbereich) in die durch den Auftraggeber gewählte Ziel-Operationalisierung einbezogen.

Grundlagen und Voranalysen

Die Test- und Befragungserhebungen bei einer Stichprobe von rund 3 800 Maturandinnen und Maturanden aus der Deutschschweiz und der Romandie und die Analysen von schriftlichen Maturaprüfungen aller drei Sprachregionen stützen sich auf folgende Grundlagen: Analyse von Lehrmaterialien und ersten Zwischenprüfungen der 16 gemessen an den Studierendenzahlen grössten universitären Studienfächer der Schweiz, Befragung der Dozierenden der Lehrveranstaltungen dieser Fächer und Einbezug der aktuellen Theorie und Forschung zur Messung kognitiver Leistungen und zur Voraussage des Studienerfolgs.

Das Hauptgewicht des Tests in Erstsprache liegt auf übergreifenden, für alle Studienrichtungen wesentlichen Sprachkompetenzen.

Der Mathematiktest enthält vor allem Aufgaben zu Inhalten, die für eine breite Anzahl von Studienfächern und nicht speziell für das Studienfach Mathematik von Bedeutung sind.

Der Biologietest hat den Charakter eines Wissenstests zu Fachinhalten, die vor allem für ein Biologie-, Biochemie- oder Medizinstudium, aber auch für andere naturwissenschaftliche Studienfächer und teilweise auch für nicht naturwissenschaftliche Studien von Bedeutung sind (z. B. für Psychologie).

Der überfachliche Fähigkeitstest (UFT) lehnt sich an die Eignungsprüfung für das Medizinstudium an und misst allgemeinere kognitive Fähigkeiten. Er wurde nur ergänzend in der Deutschschweiz durchgeführt.

Das Testinstrumentarium umfasst somit verschiedenartige Prädiktoren des Studienerfolgs und kann deshalb als ausgewogen bezeichnet werden. Interpretiert man allgemeine Studierfähigkeit dahin gehend, jedes beliebige Studium ohne Probleme aufnehmen zu können, haben die in den Testbereichen gemessenen Kompetenzen folgende Bedeutung: Erstsprachkompetenzen gehören zu den Grundlagen beinahe jedes Studienfachs. Mathematik ist wichtig für eine grosse Zahl von Fächern, in denen die mathematische Formalsprache verwendet wird. Biologisches Wissen braucht es für weniger Fächer, gehört aber zu den wesentlichen Grundlagen der obengenannten Studien. Die mit dem UFT gemessenen Fähigkeiten korrelieren nachgewiesenermassen signifikant mit dem Erfolg im Medizinstudium und damit mit jenem in allen anderen Studienrichtungen, die gleiche oder ähnliche Fähigkeiten erfordern. Die Voranalysen von EVAMAR II haben im Weiteren ergeben, dass es allgemein für jedes Fachstudium von Vorteil ist, bereits zu Beginn über Wissen und Können aus dem entsprechenden Maturitätsfach – wenn auch in unterschiedlichem Umfang – zu verfügen.

Gesamtergebnisse der Tests

Im Mittel wurden die Testaufgaben etwa zur Hälfte richtig gelöst; die Resultate waren in Erstsprache etwas besser als in Mathematik und Biologie. In Biologie sind die Ergebnisse für die Gruppe der Maturandinnen und Maturanden, die dieses Fach bereits vor einem halben Jahr oder mehr abgeschlossen hatten, klar schlechter ausgefallen, vermutlich in Folge des Vergessenseffekts. Das deutet darauf hin, dass das an Maturitätsschulen erworbene Wissen nur für kurze Zeit vollumfänglich präsent ist, und dass sowohl ein möglichst später Abschluss des Faches am Gymnasium wie auch eine rasche Aufnahme des Studiums vorteilhaft sind. Die Ergebnisse des UFT liegen nur leicht unter den Resultaten des jährlich durchgeführten Medizineignungstests. Das ist plausibel, weil bei den EVAMAR-Erhebungen auch alle jene Personen teilgenommen haben, die sich ein Medizinstudium wegen mangelnder Fähigkeiten nicht zutrauen. Dieses Ergebnis ist zudem ein wichtiger Indikator für die hohe Anstrengungsbereitschaft, welche die Maturandinnen und Maturanden bei der Bearbeitung der EVAMAR-Tests offenbar an den Tag legten. Aufgrund der Nähe der Ergebnisse des UFT zu den Resultaten, wie sie im individualbiographisch hochrelevanten Kontext der Eignungsprüfung erreicht werden, braucht die Anstrengungsbereitschaft deshalb auch den Vergleich mit jener in den gewohnten notenwirksamen Testsituationen im Gymnasium nicht zu scheuen.

Sowohl zwischen Einzelpersonen als auch zwischen ganzen Klassen gibt es grosse Unterschiede in den Ergebnissen. Diese fallen grösser aus als bei den effektiv erteilten Maturanoten. Es lässt sich somit feststellen, dass erhebliche Anteile von Maturandinnen und Maturanden in den mit den Tests erfassten Bereichen über vermutlich ungenügende Kompetenzen in mindestens einem Testbereich verfügen. Diese Erkenntnis ergibt sich aber auch schon bei einer Analyse der durch die Schulen selbst vergebenen Maturanoten, am ausgeprägtesten in Mathematik. 41.4% der Maturandinnen und Maturanden wurden im Jahre 2007 in der schriftlichen Maturaprüfung mit einer ungenügenden Note bewertet. Bei der Maturaendnote für Mathematik, welche auch die Erfahrungsnoten und die mündliche Prüfung berücksichtigt, waren es immer noch 24.4%. Es muss also davon ausgegangen werden, dass nicht alle Maturandinnen und Maturanden in der ganzen Breite über genügende Kompetenzen verfügen, um jedes beliebige Studium aufnehmen zu können, ohne zusätzlich Lücken füllen zu müssen; dies obwohl sie durch die Verleihung des Maturaausweises gemäss MAR 95 als „allgemein studierfähig“ qualifiziert wurden. Diese Erkenntnis stimmt überein mit von Dozierenden diverser Studienfächer subjektiv festgestellten Kompetenzlücken in den Bereichen Mathematik und Erstsprache. Wegen des Kompensationssystems für ungenügende Noten an Maturaprüfungen ist dieses Ergebnis der Maturitätsbildung aber kaum vermeidbar.

Ergebnisse der Tests unter strukturellen Gesichtspunkten

Die Anzahl Jahre der gymnasialen Schuldauer scheint sich auf die Testergebnisse auszuwirken. Eine Dauer von mindestens vier Jahren an einem Gymnasium geht einher mit überwiegend besseren Ergebnissen im Vergleich zu jener Ausbildungsstruktur, bei der das erste von vier Jahren in einer „gymnasialen“ Klasse an der örtlichen

Sekundarschule absolviert werden kann. Dies zeigt sich beim Mathematiktest, in leichterer Ausprägung auch bei Erstsprache und in der Deutschschweiz zusätzlich beim UFT.

Langzeitgymnasien schneiden in fast allen Testbereichen besser ab als die Kurzzeitgymnasien, allerdings nur leicht. Einzig im Biologietest ergeben sich auch grössere bzw. beinahe mittlere Differenzen zugunsten der Langzeitgymnasien, obwohl bei Letzteren die Biologie-Erfahrungsnoten im Ausmass zwar leicht, aber signifikant schlechter ausgefallen sind. Dabei dürfte es sich aber eher um eine „Notenanomalie“ oder das Ergebnis einer klassen- bzw. sozialnormorientierten Benotung handeln.

Zwischen den verschiedenen Schwerpunktfachgruppen lassen sich z. T. erhebliche Unterschiede feststellen. Die Spitzenergebnisse in Mathematik der Gruppe des Schwerpunktfachs (SPF) „Physik und Anwendungen der Mathematik“ (PAM) und in Biologie der Gruppe des SPF „Biologie und Chemie“ (BCH) waren dabei systembedingt zu erwarten. Der Spitzenwert für die Maturandinnen und Maturanden des SPF „Alte Sprachen“ im Vergleich zu jenen des SPF „Moderne Sprachen“, die beim Erstsprachtest lediglich eine mittlere Rangierung erreichten, überrascht jedoch teilweise. Auch für weitere Ungleichheiten in einigen Kompetenzbereichen gibt es keine Systemerklärungen. Die Gruppe des SPF „Alte Sprachen“ hat unter dem Aspekt der Ausgeglichenheit am besten abgeschnitten, die Maturandinnen und Maturanden der neuen Schwerpunktfächer „Musik“, „Bildnerisches Gestalten“, „Philosophie/Pädagogik/Psychologie“ (PPP) eher schlechter. Bei der Analyse der effektiv erteilten Maturanoten ergeben sich im Vergleich zu den Testergebnissen einige Rangverschiebungen, und vor allem erweisen sich die Unterschiede als geringer ausgeprägt. Es fällt aber auch auf, dass es SPF-Gruppen mit beinahe 50% („Moderne Sprachen“, „Bildnerisches Gestalten“, „Musik“) oder gar über 50% (PPP) ungenügenden Noten in der schriftlichen Mathematikprüfung gibt. Bei der Maturagesamtnote in Mathematik sind es für die beiden SPF-Gruppen „Moderne Sprachen“ und „Bildnerisches Gestalten“ immer noch rund 30% ungenügende Noten. Mit Abstand die beste Gruppe beim UFT war jene des SPF PAM. Die Gruppen der neuen Schwerpunktfächer PPP und „Bildnerisches Gestalten“ sowie in einem Bereich auch „Musik“ rangieren dagegen am Ende.

Die Unterschiede zwischen den Geschlechtern stimmen mit bisherigen Untersuchungen gut überein. Maturandinnen haben eher in Erstsprache, Maturanden klar in Mathematik und im naturwissenschaftlich ausgerichteten UFT besser abgeschnitten.

Notengebung

Die Notengebung stimmt nur teilweise mit den Testergebnissen überein. Eine erste Erklärung dafür ist jene, dass mit den Tests teilweise andere Kompetenzen gemessen wurden als durch die Maturaprüfungen. Diese Unterschiede sind bedingt durch den eingeschränkten Fokus der Untersuchung und haben nichts mit mangelnder Zielerreichung der Gymnasien zu tun. Eine zweite Deutung der Befunde setzt beim Verfahren der Notengebung an. Einerseits kann eine teilweise Anpassung der Notengebung an die Leistungsfähigkeit der Klasse (Sozialnormorientierung) vermutet werden, was im Vergleich zur Orientierung an Klassen übergreifenden Gütestandards (Kriteriumsorientierung) zu verschiedenen Notenmassstäben führt. Dennoch lassen sich andererseits auch bei den Notenvergleichen strukturelle Übereinstimmungen mit den Testergebnissen feststellen, was annehmen lässt, dass sich die Lehrerinnen und Lehrer bei der Notengebung zu einem beträchtlichen Teil auch an „objektivierten Gütestandards“ orientieren. Im Weiteren sind bei der Notengebung unerklärliche „Anomalien“ zu finden, welche wiederum eine nur teilweise Kriteriumsorientierung annehmen lassen. Dazu gehört z. B. die Erfahrungsnote in Biologie, welche in den Schwerpunktgruppen „Alte Sprachen“ und „Bildnerisches Gestalten“ höher liegt als in der Gruppe „Biologie und Chemie“. Auffällig ist schliesslich, dass schriftliche Maturaprüfungen vor allem in Mathematik, aber auch in Erstsprache zu einem hohen Anteil an ungenügenden Ergebnissen führen, diese aber jeweils durch die Note der mündlichen Prüfungen und die Erfahrungsnoten „aufgebessert“ werden.

Schriftliche Maturaprüfungen

Die schriftlichen Maturaprüfungen sind in allen untersuchten Fächern auf der einen Seite häufig anspruchsvoll, erfordern zur Lösung verschiedene kognitive Fähigkeiten (eine reine Wiedergabe von zuvor auswendig gelerntem Wissen ist nicht ausreichend) und decken Bereiche ab, die für die Studierfähigkeit von Bedeutung sind. Auf der anderen Seite konnten viele Prüfungen gefunden werden, welche diesem Bild nicht entsprechen. Insgesamt sind die Aufgabenstellungen recht heterogen.

Maturaarbeiten

Die wissenschaftspropädeutische Qualität der untersuchten Maturaarbeiten ist mehrheitlich als zufriedenstellend einzustufen. Obwohl der Beweis der Nachhaltigkeit bislang fehlt, sprechen viele Indizien dafür, dass es sich bei der Maturaarbeit um eine im Hinblick auf die Studierfähigkeit sinnvolle und ertragreiche Lern- und Prüfungsform handelt.

Grenzen der Studie EVAMAR II

An dieser Stelle soll ausdrücklich auf die Grenzen der Studie EVAMAR II hingewiesen werden. Sie sind bei der Interpretation der Ergebnisse und bei der Ableitung von Massnahmen für die künftige Gestaltung des Gymnasiums zu beachten! Grundsätzlich können nur über einige ausgewählte Aspekte der Ziele und der Zielerreichung der gymnasialen Bildung Aussagen gemacht werden. Insbesondere folgende Komponenten einer umfassenden Studierfähigkeit wurden im Projekt EVAMAR II nicht untersucht und ihr Vorhandensein bei Maturandinnen und Maturanden nicht analysiert: erstens Wissen und Können, welches zwar Teil einer breiten Allgemeinbildung ist – als Grundlage zur Lösung anspruchsvoller Aufgaben in der Gesellschaft (vertiefte Gesellschaftsreife) – und damit in der gymnasialen Bildung vermittelt werden muss, aber an der Universität nicht direkt vorausgesetzt wird; zweitens zur Studierfähigkeit gehörendes notwendiges Wissen und Können aus anderen Unterrichtsfächern als Erstsprache, Mathematik und Biologie; drittens Elemente überfachlicher Kompetenzen, die im gegebenen Projektrahmen nur schlecht erfassbar gewesen wären, dennoch aber für die erfolgreiche Bewältigung eines Studiums wichtig sind (z. B. die Fähigkeit des selbstorganisierten Lernens, der allgemeinen Selbstorganisation, des Recherchierens, der raschen Verarbeitung umfangreicher Fachliteratur usw.). Die Bedeutung solcher Faktoren liesse sich nur durch weitere, im umfassenden Konzept vorgeschlagene Untersuchungen ermitteln (Längsschnittuntersuchungen bei Studierenden), und sie wären nur durch aufwändige Assessmentverfahren zu messen. Immerhin wurden solche Faktoren aber bereits im Projekt EVAMAR I über Befragungen, also über Selbsteinschätzungen und damit annäherungsweise valide ermittelt. Im Zusatzfragebogen von EVAMAR II wurden zudem diesbezüglich ergänzende Fragen zu EVAMAR I eingestreut.

Wichtigste Schlussfolgerungen und Empfehlungen des Projektleiters

1. Die Schweizer Maturandinnen und Maturanden erzielten in den untersuchten Bereichen Ergebnisse, die für standardisierte, für eine bestimmte Population entwickelte Tests normal sind. Zu dieser Normalität gehört, dass die Aufgaben im Mittel etwa zur Hälfte richtig gelöst wurden. Insofern kann der Ausbildungsstand als zufriedenstellend bewertet werden. Auch die Gymnasien entlassen aber nicht alle ihre Maturandinnen und Maturanden mit Kompetenzen, die in der ganzen Breite als mindestens genügend eingeschätzt werden können. Das ist die Bestätigung eines Befundes, der aufgrund der effektiv an den Gymnasien vergebenen ungenügenden Noten an Maturitätsprüfungen vor allem in den Bereichen Mathematik und schriftliche Erstsprachprüfung schon bisher vermutet werden konnte. Das bedeutet aber auch, dass trotz gutem Ausbildungsstand Verbesserungen grundsätzlich möglich sind. Das ist keine aussergewöhnliche Schlussfolgerung, lässt sich eine solche doch für jede Schulstufe nach einer Leistungsmessung mit standardisierten Tests ziehen. Zwingender Handlungsbedarf würde dann bestehen, wenn die Aufgaben zu klar weniger als der Hälfte richtig gelöst und der Durchschnitt der geschätzten Personenfähigkeiten klar unter einem mittleren Wert liegen würden. Das ist bei den Schweizer Maturandinnen und Maturanden nicht der Fall.

2. Interpretiert man allgemeine Studierfähigkeit dahingehend, dass jeder Maturitätsausweis genügende Eingangskompetenzen für alle möglichen Studienfächer bescheinigen sollte, so ist dies vermutlich für einen beachtlichen Teil der Maturandinnen und Maturanden nicht gegeben. EVAMAR II stellt das zwar nur für die getesteten Bereiche fest, das Bild würde sich aber vermutlich bei Einbezug weiterer Bereiche kaum ändern. Auch dieser Befund ist nicht neu, drücken sich doch diese Lücken auch in den Maturanoten aus, ermöglicht durch das Notenkompensationssystem des MAR. Damit sind auch die (subjektiven) Klagen von Universitätsdozierenden über Lücken vor allem in Erstsprache und Mathematik sowie die teilweise erheblichen Durchfallquoten in den ersten Prüfungen zumindest teilweise erklärbar. Die Grundfrage ist deshalb, ob sich künftig eine „lückenlose Studierfähigkeit" aller Maturandinnen und Maturanden erzielen lässt. Dazu müssten sicher die Bestehensnormen an den Maturaprüfungen insofern revidiert werden, als ungenügende Noten nicht mehr zugelassen würden. Das hätte aber vermutlich zwei unerwünschte Effekte, nämlich eine Senkung der Maturitätsquote und eine teilweise Anpassung der Notengebung. Eine „lückenlose Studierfähigkeit" für alle Maturandinnen und Maturanden ist deshalb eine Fiktion. Wir empfehlen aber deshalb keineswegs die Aufgabe des Anspruchs, mit dem Maturitätsausweis die allgemeine Hochschulreife mit Anrecht auf prüfungsfreien Zugang zu den Hochschulen zu bescheinigen. Die detaillierte Begründung findet sich im Hauptbericht. Ein wichtiger Grund ist, dass die Möglichkeiten ungenügender Kompetenzbereiche in der geltenden Regelung nicht beliebig sind. Die Maturandinnen und Maturanden mit ungenügenden Leistungen müssen sich immerhin über doppelte Besserleistungen in anderen Fächern ausweisen. Ferner sind die praktischen Auswirkungen der festgestellten Kompetenzlücken zu berücksichtigen: EVAMAR I und weitere Untersuchungen haben gezeigt, dass eine Kanalisierung der Studienwahl aufgrund der eigenen Möglichkeiten und der damit verknüpften Interessen von selbst erfolgt und sich damit Lücken in gewissen Fächern nur beschränkt auf den konkreten Studienerfolg auswirken. Es ist aber im Hinblick auf eine Revision des MAR zu prüfen, ob in basalen Kompetenzen, die für viele Studienfächer wichtig sind, Bestehensnormen ohne die Möglichkeit der Kompensation ungenügender Ergebnisse für alle festgelegt werden sollten. So etwa für Erstsprach- und Mathematikkompetenzen, die ihrerseits aber nicht die ganze Breite der jeweiligen Fach-Curricula ausmachen, sondern nur die breit studienrelevanten. Es ginge also dabei nur um Teile aus Erstsprache und Mathematik und würde die weiteren, unter gymnasialen Zielsetzungen durchaus ebenso wichtigen curricularen Inhalte nicht betreffen. (Es handelt sich also nicht um eine allgemeine Forderung nach der einen Mindestnote 4 in Mathematik.) Eine solche Massnahme würde dem Anspruch der maturitären Qualifikation für die allgemeine Studierfähigkeit wieder mehr Boden verschaffen. Zudem kann vermutet werden, dass absolute Bestehensvorgaben ohne Kompensationsmöglichkeiten auch eine bessere Steuerwirkung insofern hätten, als eine vorzeitige Resignation in diesen Bereichen durch die Maturandinnen und Maturanden weniger schnell „kalkuliert" würde. Da Englisch als Wissenschaftssprache ebenfalls zu den Basiskompetenzen der Studierfähigkeit gehört – was auch von den befragten Dozierenden so gesehen wird –, sollte das Erreichen einer (noch zu bestimmenden) Standardkompetenz in Englisch obligatorisch werden.

3. Die Ergebnisse der Dozierendenbefragung in Verbindung mit der Zusatzbefragung der Maturandinnen und Maturanden bestätigen bisherige Evaluationsuntersuchungen insofern, als die Fähigkeit zu selbstständigem Lernen und Arbeiten in der Selbsteinschätzung als teilweise lückenhaft beschrieben wird. Den Empfehlungen dieser Studien ist deshalb zuzustimmen: der gezielten Förderung des selbstständigen Lernens an den Gymnasien sollte noch mehr Gewicht verliehen werden. Damit würde nicht nur eine der für das weitere Studium wesentlichen allgemeinen Kompetenzen gefördert, sondern Maturandinnen und Maturanden auch besser in die Lage versetzt, Eingangslücken, die je nach gewähltem Fachstudium individuell vorhanden sein können und wohl unvermeidbar sind, selbstständig aufzufüllen. Die festgestellten Defizite bei den Erstsprachkompetenzen könnten durch eine noch konsequentere Förderung korrekter Erstsprachanwendung in allen Fächern verkleinert werden.

4. Die verstärkte Förderung des selbstständigen Lernens und Arbeitens darf aber weder auf Kosten von Fachinhalten geschehen noch auf Kosten der Breite der Fächerung am Gymnasium. Sonst könnten wieder eher fachliche Lücken im Hinblick auf die breite Fächerung der universitären Studien entstehen, und das weitere – in EVAMAR II nicht untersuchte – wichtige Ziel des MAR (vertiefte Gesellschaftsreife durch breite Allgemeinbildung) könnte

tangiert werden. Das bedingt eine noch bessere Konzentration auf das Wesentliche in allen Fächern, z. B. auf der Grundlage einer besseren Analyse der Notwendigkeiten im Hinblick auf das Ziel der vertieften Gesellschaftsreife und durch eine bessere Absprache mit den Hochschulen über die Eingangserwartungen.

5. Im Weiteren könnte das Problem der nicht erzielbaren „lückenlosen Studierfähigkeit“ durch ein grösseres Angebot von „Ausgleichskursen“ entschärft werden, die von den Universitäten studienfachbezogen angeboten werden könnten. Beispiele sind bereits heute zu finden.

6. Die gefundenen Unterschiede zwischen den Klassen sind teils sehr gross. Die Noten zeigen grosso modo in die gleiche Richtung wie die Testergebnisse, es gibt aber auch „Anomalien“. Gymnasiallehrkräfte orientieren sich bei der Bewertung offenbar recht gut an übergreifenden Gütestandards, es müssen aber auch Anpassungen an das Leistungsvermögen von Klassen vermutet werden. Das schmälert die Aussagekraft von Maturanoten. Es ist deshalb zu überlegen, wie eine vermehrte Standardorientierung gesichert werden kann. Dabei ist von zentralen Maturaprüfungen eher abzuraten, weil diese ihrerseits mit vielen problematischen Aspekten behaftet sind.

7. Die Resultate deuten darauf hin, dass die Gymnasien mit nur dreijähriger Minimal-Dauer benachteiligt sind. Der entsprechende Passus im MAR sollte deshalb überdacht werden. Die Daten von EVAMAR II zeigen auch, dass dort, wo das Untergymnasium geführt wird und damit die Dauer des Gymnasiums noch grösser ist, das fachliche Niveau bei allen erhobenen Leistungsdaten (inkl. Maturanoten) am Schluss – im Ausmass zwar nur leicht aber signifikant – höher liegt. Das sollte in der an verschiedenen Orten laufenden Diskussion über die allfällige Abschaffung des Langzeitgymnasiums beachtet werden.

8. Die gefundenen Leistungsunterschiede zwischen den verschiedenen Schwerpunktgruppen sind teilweise erheblich. Sie zeigen sich nicht nur bei den Tests von EVAMAR II, sondern in abgeschwächter Form auch bei den von den Schulen selbst vergebenen Maturanoten. Auch wenn die praktischen Auswirkungen durch die unter Punkt 2 angesprochene, „angepasste“ Studienwahl der Maturandinnen und Maturanden nicht so gravierend sein dürften, sollte einerseits den Ursachen mittels Längsschnittanalysen nachgegangen werden. Andererseits sollten Massnahmen zur Verbesserung der Ausbildungsergebnisse in studienrelevanten Teilbereichen der Fächer Mathematik und Erstsprache für einen Teil der Maturandinnen und Maturanden geprüft werden; dies betrifft vor allem jene, die einen mit dem MAR 95 – im Vergleich zu den alten Maturitätstypen – neu eingeführten Schwerpunkt gewählt haben.

9. Die Vermutung, dass hohe Maturitätsquoten mit eher schlechteren Ergebnissen einhergehen, hat sich nicht nur bei den Tests, sondern auch teilweise bei den Noten erhärtet. Die offensichtlichen Chancenungleichheiten müssen politisch diskutiert werden.

10. Die Unterschiede zwischen den Geschlechtern folgen grösstenteils dem bereits bekannten Bild. Die Bestrebungen zur Verbesserung der Leistungen von Gymnasiastinnen in Mathematik und Naturwissenschaften sind weiterzuführen. Es sollten auch mögliche Unterschiede in weiteren Fächern, die nicht Gegenstand der Untersuchungen von EVAMAR II waren, untersucht werden.

11. Die Einführung der Maturaarbeit hat sich bewährt. Die selbstständige Bearbeitung von Projekten als wichtiges Element gymnasialen Lernens sollte unter Einbezug der in den Untersuchungen von EVAMAR II gefundenen Erfolgsfaktoren in moderater Weise ausgebaut werden.

12. Das Projekt EVAMAR II war als Querschnittsprojekt angelegt. Die Resultate werfen deshalb mehr Fragen nach Ursachen der Befunde bzw. Kausalitäten auf als zuverlässig beantwortet werden können. Diesen ungeklärten Fragen sollte mittels Längsschnittanalysen unter systematischer Ausnützung aller im Verlaufe der gymnasialen Ausbildung und des späteren Studiums anfallenden Personendaten nachgegangen werden.

I Einleitung zum Gesamtprojekt

1 Reform der Maturitätsbildung

Das Schweizer Gymnasium wurde 1995 grundlegend reformiert und im Maturitätsanerkennungsreglement (MAR) (EDK, 1995) neu kodifiziert. Die Schweizer Maturität vermittelt den allgemeinen und schrankenlosen Zugang zu allen universitären Studien. Ein Numerus clausus existiert nur für das Medizinstudium, für dessen Zugang ergänzend zur maturitären Qualifikation ein Eignungstest (Hänsgen & Spicher, 2007) absolviert werden muss. Mit der Reform sollte die Maturitätsausbildung neuen Anforderungen angepasst und gleichzeitig der freie Zugang zu allen Universitäts- und ETH-Studien weiterhin gewährleistet bleiben.

Bis zur Reform 95 konnte am Schweizer Gymnasium aus den fünf Maturitätstypen A bis E der zur individuellen Eignung und Neigung passende gewählt werden. Die Kernidee der Reform 95 war, an die Stelle dieser Maturitätstypen die Einheitsmatura zu setzen und gleichzeitig breite Wahl- bzw. Individualisierungsmöglichkeiten offen zu lassen.

Folgende Hauptmerkmale kennzeichnen die Einheitsmatura:

- Es gibt eine relativ hohe Zahl an im Wesentlichen für alle Gymnasiastinnen und Gymnasiasten gleichen Grundlagenfächern, um eine breit gefächerte und ausgewogene Ausbildung sicherzustellen. Es handelt sich um folgende Fächer: Erstsprache, eine zweite Landessprache, eine dritte Sprache (d. h. eine dritte Landessprache, Englisch oder eine alte Sprache), Mathematik, Naturwissenschaften mit obligatorischem Unterricht in Biologie, Chemie und Physik, Geistes- und Sozialwissenschaften mit obligatorischem Unterricht in Geschichte und Geographie sowie einer Einführung in Wirtschaft und Recht, Bildnerisches Gestalten und/oder Musik.
- Alle Lernenden müssen zusätzlich ein Schwerpunktfach und ein Ergänzungsfach besuchen, und es ist eine Maturaarbeit zu erstellen.

Folgende Hauptmerkmale charakterisieren die Individualisierung:

- Das Schwerpunktfach kann aus acht Varianten gewählt werden: alte Sprachen (Latein und/oder Griechisch), eine moderne Sprache (eine dritte Landessprache, Englisch, Spanisch oder Russisch), Physik und Anwendungen der Mathematik, Biologie und Chemie, Wirtschaft und Recht, Philosophie/Pädagogik/Psychologie, Bildnerisches Gestalten, Musik.
- Das Ergänzungsfach kann aus 13 Varianten gewählt werden: Physik, Chemie, Biologie, Anwendungen der Mathematik, Geschichte, Geographie, Philosophie, Religionslehre, Wirtschaft und Recht, Pädagogik/Psychologie, Bildnerisches Gestalten, Musik, Sport.
- Das Thema für die Maturaarbeit kann individuell in Absprache mit der betreuenden Lehrperson bestimmt werden und auch interdisziplinär ausgerichtet sein.

Aus den weiteren Vorgaben des MAR (EDK, 1995) seien folgende genannt:

- Das Maturitätszeugnis darf frühestens nach 12 Jahren Schuldauer ausgestellt werden, wovon vier Jahre an einer Maturitätsschule absolviert sein müssen. Ausnahmsweise sind auch drei Jahre Gymnasium ausreichend, falls auf der Sekundarstufe I eine „gymnasiale Vorbildung“ erfolgt ist; diese Regelung ist in gymnasialen Kreisen umstritten.
- Für das Bestehen der Maturaprüfungen darf die doppelte Summe aller Notenabweichungen von 4 nach unten nicht grösser als die Summe aller Notenabweichungen von 4 nach oben sein. Um also die ungenügende Note 3 kompensieren zu können, braucht es z. B. die Note 5 in zwei anderen Fächern.

Es gibt für die Schweizer Gymnasien weder verbindliche Bildungs- oder Ausbildungsstandards noch eine zentrale Prüfung der Zielerreichung. Inhaltliche Vorgaben sind lediglich im Zielartikel des MAR (EDK, 1995) und im Rahmenlehrplan (RLP) (EDK, 1994) der Schweizerischen Konferenz der kantonalen Erziehungsdirektoren (EDK) zu finden. Die wesentlichen Inhalte des Bildungszielartikels sind in den folgenden zwei Sätzen zusammengefasst (EDK, 1995, Art. 5): „Die Schulen streben eine breit gefächerte, ausgewogene und kohärente Bildung an, nicht aber eine fachspezifische oder berufliche Ausbildung. Schülerinnen und Schüler gelangen zu jener persönlichen Reife, die Voraussetzung für ein Hochschulstudium ist und die sie auf anspruchsvolle Aufgaben in der Gesellschaft vorbereitet." Der zweite Satz beinhaltet dabei die zwei obersten Bildungsziele des Gymnasiums, nämlich die Erlangung der Hochschulreife und die Vorbereitung auf anspruchsvolle Aufgaben in der Gesellschaft, der erste Satz verweist auf den grundsätzlichen Weg zur Zielerreichung (siehe zu Überlegungen zur Stimmigkeit und zur wissenschaftlichen Richtigkeit und Kohärenz des ganzen Bildungszielartikels Eberle, 2008).

Im RPL sind in etwas differenzierterer, aber immer noch wenig operationalisierter Form sowohl die fachübergreifenden als auch die Ziele der einzelnen Fächer beschrieben. Bemerkenswert sind Erstere, die sog. „überfachlichen Kompetenzziele", welche folgende Komponenten umfassen (EDK, 1994, 11):

- den sozialen, ethischen und politischen Bereich,
- den intellektuellen, wissenschaftlichen und erkenntnistheoretischen Bereich,
- den kommunikativen, kulturellen und ästhetischen Bereich,
- Bereiche der Persönlichkeitsentwicklung und der Gesundheit, sowie
- Bereiche der persönlichen Lern- und Arbeitstechniken, der Wissensbeschaffung und der Informationstechnologien.

Dabei wird zwischen Grundkompetenzen für alle Jugendlichen und ergänzenden Kompetenzen speziell von Gymnasiastinnen und Gymnasiasten unterschieden.

Die Darstellung der Fachziele gliedert sich für jedes Fach in „Allgemeine Bildungsziele", „Begründungen und Erläuterungen" und „Richtziele". Die Richtziele sind wiederum unterteilt in „Grundkenntnisse", „Grundfertigkeiten" und „Grundhaltungen". Auf die Ziele im Einzelnen und die vorhandenen Schwächen des RLP – insbesondere mangelnde Theorieabstützung der überfachlichen Kompetenzziele und dürftige Fundierung sowie Kohärenz der fachlichen Ziele – kann an dieser Stelle nicht eingegangen werden.

Bei den beschriebenen Grundsätzen handelt es sich nur um Rahmenvorgaben; und die Stundenanteile für die einzelnen Fächer sind lediglich durch Prozentbänder im MAR 95 abgesteckt. Der gesamtschweizerische Lehrplan wird ausdrücklich als Rahmenlehrplan bezeichnet, alle 26 Kantone und Halbkantone haben ihre jeweils eigenen Umsetzungen des RLP vorgenommen. In grossen Kantonen, wie z. B. dem Kanton Zürich, verfügen sogar die einzelnen Gymnasien über ihre eigenen Lehrpläne. Dieser Programmvielfalt entsprechen auch die Maturitätsprüfungen: Weder auf Bundes- noch auf Kantonsebene werden einheitliche Prüfungen durchgeführt. Es ist durchaus üblich, dass parallel im gleichen Fach unterrichtende Lehrkräfte an der gleichen Schule ihre je eigene Maturitätsprüfung zusammenstellen und die Notenmassstäbe selbst festlegen. Trotz dieser potenziellen Vielfalt der daraus resultierenden Qualifikations- oder Reifemuster der Schweizer Maturandinnen und Maturanden gilt eine solche „lokale" Matura als die Schweizer Matura, welche den Zugang zu allen Schweizer Universitäten eröffnet.

An den Schweizer Gymnasien werden also auch nach der Maturitätsreform 95 explizit von zentraler Stelle nur lose verbindliche Ziele gesetzt. Das bedeutet aber nicht, dass Standards nicht zumindest implizit dezentral oder lokal vorhanden wären (siehe auch Oelkers, 2006):

- Lehrkräfte haben in der Regel ein professionelles, sich an fachspezifischen Standards orientierendes Fachverständnis.
- Lehrmittel wirken standardisierend; allerdings gibt es keine einheitlichen Lehrmittel.
- Explizite oder implizite Aufnahmequoten an die Gymnasien setzen Standards bezüglich Voraussetzungen der Schülerinnen und Schüler.

Es bleiben aber offensichtliche Unterschiede zwischen Kantonen und Schulen. Erwähnt seien an dieser Stelle nur diese:

- Zum Teil differieren die Stundenzahlen für einzelne Fächer erheblich.
- Die Maturitätsquoten in den Kantonen (Anteil der Jugendlichen, welche eine gymnasiale Matura erwerben) reichten beim Maturajahrgang 2007 von 11.2% im Kanton Obwalden bis zu 29.5% im Kanton Tessin; der gesamtschweizerische Durchschnitt lag bei 19.2% (Bundesamt für Statistik, 2008).

Das Maturitätszeugnis sollte Auskunft geben über die am Gymnasium erworbenen Qualifikationen und damit auch über den Grad des Erreichens des Maturitätsziels der Hochschulreife. Sein prognostischer Wert wird in den letzten Jahren von der abnehmenden Institution Universität vermehrt in Frage gestellt, was sich z. B. an der Verwendung eigener Testverfahren beim Numerus-clausus-Fach Medizin und der Einführung von strengen Selektionsverfahren im Rahmen von Assessment-Jahren zeigt. Die oben erwähnten Mängel dürften dazu beitragen. Die in Deutschland gefundenen erheblichen Zusammenhänge zwischen Abitur- und Prüfungsnoten an Universitäten von bis zu rho=.53 (siehe die Metaanalyse bei Trapman et al., 2007) dürfen nicht ohne Weiteres auf die Schweiz übertragen werden, weil die in den deutschen Untersuchungen verwendeten Noten standardisierter sind. Neue, repräsentative Schweizer Daten fehlen leider. Eine Ausnahme wird eine Studie „Maturanoten und Studienerfolg“ der ETH Zürich sein, die in nächster Zeit veröffentlicht wird.

Bis im Sommer 2001 musste die Reform 95 in den Kantonen umgesetzt sein, direkt im Anschluss begann die Evaluation der Reform. Neben vielen kantonalen Untersuchungen wurde durch die für die Gesamtschweiz zuständigen politischen Behörden (Bund und EDK) das Projekt EVAMAR (Evaluation der Maturitätsreform 95) lanciert. Der erste Teil (EVAMAR I) war Ende 2004 abgeschlossen. Im Juni 2005 beschlossen EDK und Bund den Beginn einer zweiten Evaluationsphase (EVAMAR II), die bis Ende 2008 läuft.

2 EVAMAR I

In Phase I von EVAMAR wurden im Wesentlichen die folgenden drei Themen untersucht (Ramseier et al., 2005): (1) Die Passung von Wahlfachangebot und Interessen der Schülerinnen und Schüler sowie die Auswirkungen auf den Ausbildungserfolg, insbesondere die subjektiv wahrgenommene Qualität der Vorbereitung auf ein Hochschulstudium; (2) die Umsetzung der fächerübergreifenden pädagogischen Ziele; (3) die Bewältigung der Reformen durch die Schulen. Hauptinstrument war die Befragung von rund 10'000 Maturandinnen und Maturanden, von Lehrpersonen und von Schulleitungsmitgliedern. Leistungsmessungen wurden nicht eingesetzt. Die Ergebnisse von EVAMAR I liegen seit Ende 2004 vor. Die wichtigsten Resultate der ersten beiden Analysen sind die folgenden (Ramseier et al. 2005):

(1) Ergebnisse der Selbsteinschätzungen der Maturandinnen und Maturanden zur Passung von Wahlfachangebot und Interessen der Schülerinnen und Schüler sowie die Auswirkungen auf den Ausbildungserfolg, insbesondere die subjektiv wahrgenommene Qualität der Vorbereitung auf ein Hochschulstudium:

- Es ist eine Zunahme der Individualisierung der gymnasialen Ausbildung und eine Anpassung an die Interessen der Schülerinnen und Schüler erfolgt. Die alten Sprachen haben dabei an Bedeutung verloren.
- Die Wahlmöglichkeiten und die Abdeckung der Fachinteressen erhöhen die Lernmotivation.
- Die generelle Selbsteinschätzung der Lernenden hinsichtlich der Vorbereitung auf die Hochschule ist gut. Aus ihrer Sicht wird somit das Ziel der allgemeinen Hochschulreife erreicht.
- Der Zusammenhang zwischen Bildungsprofil (gewähltes Schwerpunkt- und Ergänzungsfach) und Studium ist eher locker, noch am engsten findet er sich bei Wirtschaft, Mathematik/Informatik und „harten“ Ingenieurwissenschaften.

- Die Misserfolgsquote bei den Maturaprüfungen hat in der Deutschschweiz zugenommen. Mathematik stellt die grösste Hürde unter den einzelnen Fächern dar.
- Die Zahl der erfolgreichen Gymnasiastinnen ist höher als jene der Gymnasiasten, jedoch planen mehr männliche Gymnasiasten ein Hochschulstudium als weibliche.

(2) Ergebnisse der Einschätzungen zur Umsetzung der fächerübergreifenden pädagogischen Ziele:

- Bei den Maturaarbeiten überwiegen in der Deutschschweiz empirische Arbeiten, in der übrigen Schweiz theoretische. Aus Sicht der Lehrpersonen werden das Fehlen befriedigender Beurteilungskriterien und der Zeitmangel kritisch bewertet. Lernende erachten die Betreuungsqualität und die erhaltenen Anweisungen als Knackpunkte.
- Der fächerübergreifende Unterricht erfolgt hauptsächlich durch den „transdisziplinären Blick" vom eigenen Fach auf andere, also in der Regel ohne eigene Zeitgefässe; er wird insgesamt positiv beurteilt. Vorbehalte gab es nur bei Mathematik und Naturwissenschaften, als Hauptschwierigkeit wurde hier Zeitnot angeführt.
- Die Förderung der fächerübergreifenden Kompetenzen wird unterschiedlich erfolgreich eingeschätzt, von Lehrenden zumeist positiver als von Schülerinnen und Schülern.

Insgesamt wurde die Reform überwiegend als zufriedenstellend beurteilt.

Im Hinblick auf eine Annäherung an das übergeordnete Interesse eines gut begründeten und kohärent umgesetzten Systems der Maturitätsbildung weisen wir in diesem Abschnitt auf eine in den Untersuchungen der Phase I nach unserer Einschätzung offen gelassene Frage hin. Als die entscheidenden Neuerungen des MAR 95 (1995) gegenüber der vorangegangenen Regelung durch die Maturitätsanerkennungsverordnung MAV 68 (1968) werden im Projektbericht die folgenden Elemente genannt: der Ersatz des Typensystems durch ein System von Grundlagen-, Schwerpunkt- und Ergänzungsfächern sowie die starke Gewichtung des fächerübergreifenden Unterrichts und der Förderung fächerübergreifender Kompetenzen durch die Einführung der Maturaarbeit und durch die Gruppierung der naturwissenschaftlichen bzw. der geistes- und sozialwissenschaftlichen Fächer zu je einem Grundlagenfach. Diese Auswahl übersieht zumindest eine weitere wesentliche Innovation: Die geistes- und sozialwissenschaftlichen Fächer wurden nicht nur zusammengefasst, sondern vor der Zusammenfassung um ein weiteres Unterrichtsfach ergänzt, nämlich Wirtschaft und Recht. Die Einführung eines neuen Faches mit den entsprechenden Folgemassnahmen (Stundentafeln, neue Curricula, Ausbildung von Lehrkräften usw.) müsste unbedingt ebenfalls in die Evaluation einbezogen werden, auch wenn man in einer restriktiven Interpretation lediglich von einer Erweiterung der gymnasialen Bildungsinhalte im Grundlagenfach „Geistes- und Sozialwissenschaften" sprechen könnte. Faktisch sind nämlich die Einzelfächer geblieben, und deshalb gibt es ein neues gymnasiales Fach. Evaluationsbedürftig wären im Weiteren auch die Umsetzung und die Wirkung des neuen Ansatzes für den Einbezug der Informatik in die gymnasiale Bildung, wonach dieses Gebiet kein eigenes Fach mehr ist, sondern in andere Fächer integriert werden sollte.

3 Weitere Evaluationsstudien

Neben den vielen kantonalen Evaluationsstudien sollen an dieser Stelle die Hauptergebnisse der ebenfalls auf gesamtschweizerischer Ebene durchgeführten zwei Studien von Notter & Arnold (2003 und 2006) dargestellt werden. Diese erforschten ähnliche Fragen wie EVAMAR I, die Befragung wurde aber nicht bei Gymnasiastinnen und Gymnasiasten, sondern im Januar 2002 und im Januar 2005 bei je einer Kohorte von Studierenden an Schweizer Universitäten im dritten Semester durchgeführt. Dabei hatte die erste Kohorte das Maturazeugnis noch mehrheitlich nach alter, die zweite hingegen nach neuer Ordnung erworben. Die wichtigsten, vergleichenden Ergebnisse der zweiten Studie sind die folgenden (Notter & Arnold, 2006, S. 7 ff.):

- „Die Studierenden schätzen ihr an der Mittelschule erworbenes Wissen und Können in den Bereichen der Muttersprache (Grammatik, Orthographie, Literatur etc.) als gut bis eher sehr gut ein.“ Ihr diesbezügliches Wissen und Können schätzen die Absolventen aller Schwerpunktfächer und in allen Sprachregionen im Vergleich zu den anderen Fächern als am höchsten ein, ebenso wie sie es in vielen Studienfachbereichen als mindestens „eher wichtig“ für das Studium beurteilen.
- „Die Studierenden nach neuer und alter Maturitätsordnung schätzen ihre Fachkenntnisse in den meisten Fächern sehr ähnlich ein. Nur in Informatik-Anwendungen schätzen die Studierenden nach neuer Maturitätsordnung ihre Fachkenntnisse deutlich höher und in alten Sprachen sowie in geringem Masse auch in Physik tiefer ein als die Studierenden nach alter Maturitätsordnung.“
- „Die Studierenden nach neuer Maturitätsordnung schätzen ihre überfachlichen Kompetenzen meistens höher ein als die Studierenden nach alter Maturitätsordnung.“
- „Auch ihre Kompetenzen im Lernen mit verschiedenen Lehr- und Lernmethoden schätzen die Studierenden nach neuer Maturitätsordnung meistens höher ein als die Studierenden nach alter Maturitätsordnung. Dies ist besonders beim Lernen durch längerfristige selbständige Arbeiten (Maturaarbeit) der Fall.“
- „Die Studierenden nach neuer Maturitätsordnung sind im Durchschnitt, wie schon die Studierenden nach alter Maturitätsordnung, zufrieden bis sehr zufrieden mit „ihrer“ Mittelschule in Bezug auf die Vorbereitung auf ihr Studium. In Bezug auf ihre persönliche Entwicklung sind Studierende nach neuer Maturitätsordnung etwas zufriedener, was wohl auf die grössere Wahlfreiheit zurückzuführen ist.“
- „Im Allgemeinen scheint die Passung zwischen dem an der Mittelschule erworbenen Wissen und Können und den Anforderungen des Studiums gut zu sein. Im Bereich des fachlichen Wissens gilt dies primär für diejenigen Schwerpunktfächer, die für diesen Studienfachbereich typisch sind. Bei den für das jeweilige Studienfach atypischen Schwerpunktfächern zeigt sich dagegen oft eine Diskrepanz zwischen der Einschätzung des eigenen Könnens zum Zeitpunkt der Matura und der Wichtigkeit der für diesen Fachbereich bezeichnenden Fächer“, indem die Letztere höher liegt.
- „Der Vergleich der Einschätzung des eigenen Könnens zum Zeitpunkt der Maturität und der Wichtigkeit dieses Könnens für das erste Studienjahr zeigt auch gewisse Diskrepanzen auf. So beanspruchen in den meisten Studienfachbereichen Informatik-Anwenderkenntnisse und z. T. auch Informatik-Programmierkenntnisse eine gewisse Wichtigkeit, die Studierenden schätzen allerdings ihre diesbezüglichen Kenntnisse zum Zeitpunkt der Maturität immer noch deutlich tiefer ein, obwohl sie ihre diesbezüglichen Kenntnisse im Vergleich zu Studierenden nach alter Maturitätsordnung höher einschätzen.“
- In allen Studienfachbereichen wird lediglich den Englischkenntnissen grosse Bedeutung beigemessen, anderen Fremdsprachenkenntnissen nicht.

Die Autoren kommen insgesamt zum Schluss, dass „es den Mittelschulen auch mit der neuen Maturitätsordnung gut zu gelingen“ scheint, „den Spagat zwischen dem Vermitteln einer breiten Allgemeinbildung und dem Vermitteln von spezifischen Fachkenntnissen als Vorbereitung für ein universitäres Fachstudium zu machen.“

4 Teilrevision des MAR 95

Weniger aufgrund der Ergebnisse der verschiedenen Evaluationsuntersuchungen, sondern vor allem auf Drängen der gymnasialen Standesorganisationen ist auf das Jahr 2008 hin eine Minireform des MAR 95 beschlossen worden, die folgende wesentliche Elemente umfasst (EDK, 2007):

- Die naturwissenschaftlichen Fächer (Biologie, Physik und Chemie) werden nicht mehr als Fächergruppe (eine Note), sondern als Einzelfächer (drei Noten) bewertet.
- Der Unterrichtsanteil für die mathematisch-naturwissenschaftlichen Fächer wird auf 25-35% (heute 20-30%) erhöht.

- Informatik wird neu in den Katalog der Ergänzungsfächer aufgenommen.
- Die unter „Geistes- und Sozialwissenschaften“ zusammengefassten Fächer Geschichte und Geographie werden einzeln aufgeführt und mit je einer eigenen Note ausgewiesen. Die bisher ebenfalls in dieser Fächergruppe aufgeführte Einführung in Wirtschaft und Recht wird zwar nicht benotet, bleibt weiterhin obligatorisch.
- Die Kantone haben neu die Möglichkeit, Philosophie in den Katalog der Grundlagenfächer aufzunehmen.
- Der Stellenwert der Maturaarbeit wird dadurch erhöht, dass ihre Benotung künftig für das Bestehen der Maturaprüfung zählt.

5 Auftrag und Fragestellung von EVAMAR II

Im Sommer 2005 beschlossen Bund und EDK den Beginn der zweiten Evaluationsphase. Das Projekt EVAMAR II zielte – in Ergänzung zu EVAMAR I und ähnlichen Studien – auf die Untersuchung weiterer Aspekte der Maturitätsbildung. Das Schwergewicht sollte auf der objektivierten Erfassung des Ausbildungsstandes der Schülerinnen und Schüler am Ende des Gymnasiums liegen und Leistungsmessungen miteinschliessen. Die in EVAMAR II vorgesehenen Untersuchungen wurden aus einem umfassenden Evaluationskonzept ausgewählt, das die Projektbearbeiter im Auftrag von Bund und EDK bereits im Sommer 2004 entwickelt hatten. Der Auftrag zur Projektleitung von EVAMAR II ging in der Folge ebenfalls an das Institut für Gymnasial- und Berufspädagogik (IGB, vormals Höheres Lehramt Mittelschulen) der Universität Zürich.

Im Folgenden werden zum besseren Verständnis von EVAMAR II zuerst die Grundzüge des erwähnten umfassenden Konzepts zur Evaluation der Maturitätsbildung beschrieben (Abschnitt 5.1). In Abschnitt 5.2 schliesst sich eine Kurzbeschreibung der für EVAMAR II ausgewählten Untersuchungen an.

5.1 Das umfassende Untersuchungskonzept

Die ideale Evaluation einer Reform würde darin bestehen, im Sinne einer Kausalanalyse die Wirkung der Gesamtreform und der einzelnen, auch kantonal gestalteten Reformmassnahmen auf der Ebene der Schulen und der Auszubildenden über einen Vorher-Nachher-Vergleich zu erfassen. Das war hier nicht möglich. Bereits die Autoren des Berichts über EVAMAR I haben festgehalten, dass „die grosse Vielfalt der kantonalen Reformmassnahmen – seien es die konkrete Umsetzung der schweizerischen Vorgaben oder die zusätzlich durchgeführten Reformmassnahmen ...“ einen nationalen Vorher-Nachher-Vergleich verunmöglicht und der „Evaluationsanspruch entsprechend eingegrenzt“ werden muss (Ramseier et al., 2005, 38f.). Dem ist zuzustimmen. Zudem könnte auch ein Vorher-Nachher-Vergleich vorwiegend nur ex-post erfolgen und wäre forschungsmethodisch einerseits wegen weitgehend fehlender standardisierter und valider Vorher-Daten und andererseits wegen des Einflusses unkontrollierbarer Variablen während der Umstellungsphase problematisch (z. B. Bortz & Döring, 2002, 518 ff.) oder zumindest äusserst aufwändig. Die Untersuchung musste deshalb von vornherein als Querschnitt angelegt werden und hat eher den Charakter der Erfassung des Ist-Zustandes, dessen Ursachen nicht allein auf die Reform zurückgeführt werden können. Im Hinblick auf Massnahmen zur Weiterentwicklung des Gymnasiums verspricht diese Untersuchungsanlage aber gleichwohl ein erhebliches Erkenntnispotential.

In der Vorphase der Auftragserteilung des Projekts EVAMAR II wurde auf der Grundlage eines Konzepts zur umfassenden Evaluation der Maturitätsbildung eine breite Palette verschiedener, miteinander zusammenhängender Untersuchungen unter forschungsmethodischen sowie pragmatischen Aspekten diskutiert, und sie wurden unter der Prämisse beschränkter Ressourcen auch gegeneinander abgewogen sowie mit unterschiedlichen Prioritäten versehen. Wesentliche Elemente des umfassenden Konzepts sind die nachstehend beschriebenen:

Eine evaluative Erfassung des Ausbildungsstandes muss sich sinnvollerweise an den Zielen der Ausbildung orientieren. Die Ziele für Maturitätsschulen werden einerseits – zumindest in allgemeiner Form – durch die gesetzlichen Grundlagen des Gymnasiums vorgegeben und andererseits in der Literatur diskutiert (z. B. Huber, 1997 und 1998). Die Hauptziele der gymnasialen Bildung sind gemäss Art. 5 des MAR 95 (1995) die Erlangung „jener persönlichen

Reife“, die erstens „Voraussetzung für ein Hochschulstudium ist“ (Hochschulreife oder Studierfähigkeit), und die zweitens „auf anspruchsvolle Aufgaben in der Gesellschaft vorbereitet“ (vertiefte Gesellschaftsreife durch breite Allgemeinbildung).[1] Im Projektauftrag und in der Diskussion um das Gymnasium wird anstelle der „Hochschulreife“ auch der Begriff „Studierfähigkeit“ verwendet. Auch wenn sich in einer vertieften Diskussion neben Gemeinsamkeiten durchaus Unterschiede in den Bedeutungen herausschälen liessen, wurden im Rahmen des Projekts EVAMAR II die Begriffe „Studierfähigkeit“ und „Hochschulreife“ synonym interpretiert im Sinne aller Kompetenzen (Kenntnisse, Fähigkeiten, Fertigkeiten und Bereitschaften), die unabdingbare Voraussetzung für die erfolgreiche Bewältigung eines universitären Hochschulstudiums sind, also befähigen, ein solches erfolgreich zu beginnen, durchzuführen und abzuschliessen (vgl. auch Huber, im Druck). Allgemeine Studierfähigkeit oder Hochschulreife meint dann, dass sich diese Kompetenzen in Übereinstimmung mit der allgemeinen Hochschulzugangsberechtigung auf jedes Hochschulstudium in jedem Fach erstrecken. Die Beziehung zwischen diesem ersten Teilziel der Studierfähigkeit und dem zweiten Teilziel der Vorbereitung auf anspruchsvolle Aufgaben in der Gesellschaft ist nicht eindeutig. Die Interpretationen schwanken zwischen einer Hierarchie der Ziele (Erreichung des primären Zieles der Hochschulreife ist übergeordnet und dient auch dem zweiten Teilziel) (z. B. Szidat, 2003) und einer Gleichordnung relativ unabhängiger Ziele (z. B. Ernst, 2004). Die gleichgeordnete Interpretation scheint uns konsistenter. Die beiden Ziele haben teils gleichgerichtete, teils aber auch eigenständige und mitunter breite curriculare Auswirkungen. Die weiteren Zielbeschreibungen in Art. 5 MAR 95 (1995) lassen sich logisch als Komponenten der beiden Hauptziele einordnen. Z. B. dient die „Einsicht in die Methodik wissenschaftlicher Arbeit“ der Erlangung der Hochschulreife, oder die Fähigkeit, „sich den Zugang zu neuem Wissen zu erschliessen“, ist (auch im Sinne Lebenslangen Lernens) sowohl für die Studierfähigkeit wie auch für die Lösung anspruchsvoller Aufgaben von grosser Bedeutung (siehe im Einzelnen bei Eberle, 2008).

Eine vollständige und an fundierten Kriterien orientierte Evaluation muss somit zunächst die genügende Operationalisierung der sich überschneidenden Ziele der Studierfähigkeit und der breiten Allgemeinbildung umfassen. In der Schweiz gibt es keine Untersuchung, in der diese Arbeit schon geleistet worden wäre. In Deutschland liegt eine solche, auf einer soliden empirischen Basis stehende Studie ebenfalls nicht vor (Huber, 1998, 154 f. und Huber im Druck). Das Projekt EVAMAR II bot nun die Gelegenheit, die Einlösung dieses längst fälligen Desiderates anzuregen. Die Überlegungen zur Konkretisierung der Maturitätsziele sind in Abbildung I_1 dargestellt. Bei den angegebenen Untersuchungsmethoden handelt es sich um eine nicht abschliessende Auswahl von Verfahren, mit denen die zurzeit realen Anforderungen ermittelt werden könnten. Zu den Untersuchungsmethoden gehören:

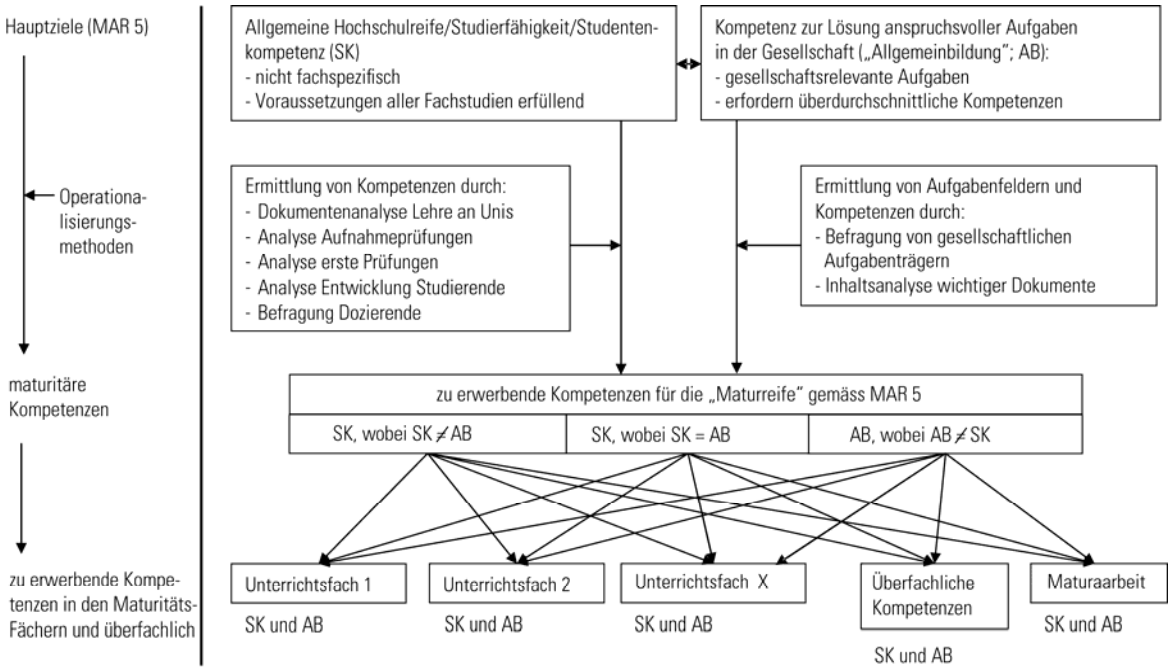
- Klärung des Konstrukts „Hochschulreife“ bzw. „Studierfähigkeit“ aus Sicht der Hochschulen durch:
  - Inhaltsanalyse der Fachcurricula, Skripten, zentralen Lehrbücher usw. an Universitäten; Ermittlung der vorausgesetzten fachlichen Anforderungen der verschiedenen Studienfächer.
  - Inhaltsanalyse von Aufnahmeprüfungen; Ermittlung der kognitiven Anforderungen.
  - Inhaltsanalyse der ersten Prüfungen an den Universitäten (z. B. Zwischenprüfungen; Assessment-Stufen-Prüfungen; Prüfungen im ersten Semester ETH); Ermittlung der kognitiven Anforderungen.
  - Ermittlung der benötigten allgemeinen Studierkompetenzen durch Analyse der Lern- und Denkprozesse von Studierenden; Längsschnittuntersuchungen (Fallstudien): Anleitung zur Führung eines Lernjournals mit anschliessender Inhaltsanalyse; mündliche Interviews, „Lautes Denken“.
  - Befragung von Lehrstuhlinhabern, Dozierenden von Grundlagenvorlesungen und Studienberatern der einzelnen Fakultäten über Studienanforderungen und -voraussetzungen.
- Operationalisierung der „anspruchsvollen Aufgaben“ und Ableitung der zu ihrer Lösung erforderlichen Kompetenzen als Grundlage der durch breite Allgemeinbildung erlangten vertieften Gesellschaftsreife durch:
  - Befragung von und vertiefende Interviews mit gesellschaftlichen Aufgabenträgern.

[1] Dieser zweite Teil wird allerdings in der einschlägigen Diskussion nicht immer explizit aufgenommen (z. B. Oelkers, 2008).

- Inhaltsanalyse von Dokumenten mit normativen Vorgaben für die gesellschaftliche Entwicklung (z. B. Agenda 21 für eine nachhaltige Entwicklung); Analyse vorhandener Kompetenzmodelle mit Anforderungsprofilen für Führungskräfte von grossen Unternehmungen; Analyse statistischer Erhebungen der späteren Berufskarriere von Maturandinnen und Maturanden.

Aus den Ergebnissen dieser Untersuchungen liessen sich für jedes Maturitätsfach und für überfachliche Bereiche die entsprechenden durch die Lernenden zu erwerbenden, zieladäquaten Kompetenzen konkretisieren. Damit wären die Kriterien für eine umfassende Evaluation der Maturitätsbildung ausreichend genau bestimmt. Auf die methodischen und normativen Probleme dieser Untersuchungen kann an dieser Stelle nicht eingegangen werden.

Abb. I.1: Umfassendes Modell zur Herleitung der Evaluationskriterien



Legende: - AB: Abkürzung für „Allgemeinbildung"
- SK: Abkürzung für „Studierkompetenz"

Die Klärung der zu erwerbenden Zielkompetenzen bzw. die Bestimmung der Evaluationskriterien im beschriebenen ersten Evaluationsschritt ist die Grundlage für deren Anwendung im zweiten Teil des Evaluationsvorhabens. Die Erreichung der Zielgrössen kann über verschiedenste Indikatoren untersucht werden. Die nachfolgende Aufzählung ist nicht abschliessend:

- Ergebnisse der Maturitätsbildung: Entwicklung und Durchführung von Tests in allen Fächern mit Maturandinnen und Maturanden zur Überprüfung der operationalisierten Zielkompetenzen.
- Qualität der Lehrpläne und der Stundentafeln: Inhaltsanalyse der Lehrpläne und Stundentafeln und Vergleich mit den Zielkompetenzen.
- Qualität des Unterrichts: Unterrichtsevaluation durch Inhaltsanalyse von Unterrichtsdokumenten und mündlichen Interviews; evtl. ausgewählte Video-Analysen.
- Qualität der Maturitätsprüfungen:
  - Inhaltsanalyse der schriftlichen Maturitätsprüfungen (insbesondere fachinhaltlich und bezüglich kognitivem Anspruchsniveau) und Vergleich der Ergebnisse mit den in den einzelnen Fächern zu erwerbenden Zielkompetenzen sowie mit den Ergebnissen der durchgeführten Tests (siehe oben); Video-Analysen mündlicher Prüfungen.

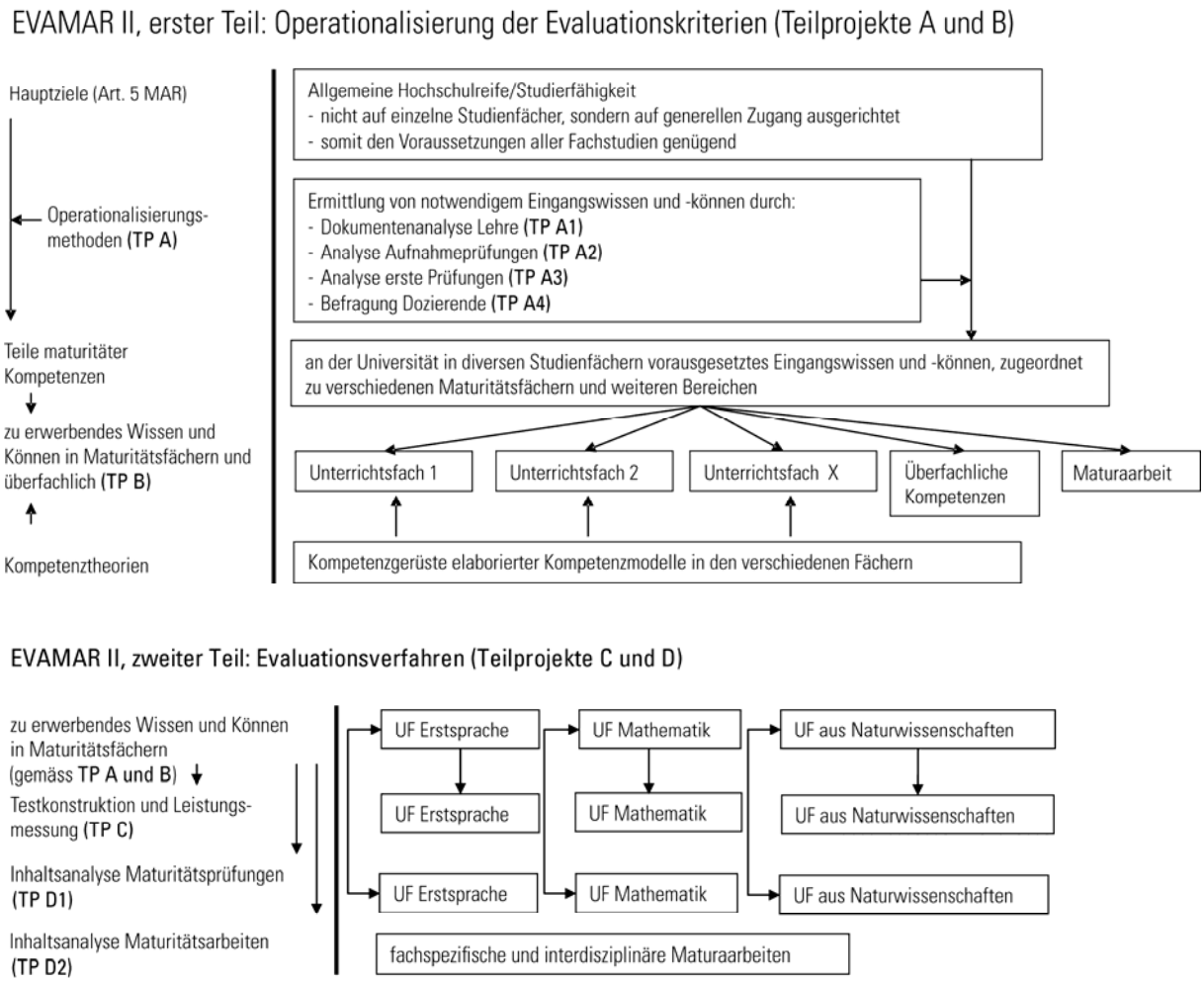
- Inhaltsanalyse der Maturaarbeiten in Bezug auf Themen- und Fächerverteilung, Arbeitsprozess der Lernenden, Form und Qualität der Arbeit, Bezug zu grundlegenden, wissenschaftlichen Arbeitstechniken wie Quellenstudium und deren korrekte Verarbeitung, Beurteilung der Arbeit, Vergleichbarkeit u. a.; Befragungen zu Rolle und Funktion der Betreuer.
- Vergleich der Noten an Maturitätsprüfungen mit den Ergebnissen von Prüfungen an den Universitäten (statistische Analysen).

Bei diesen Untersuchungen sollte explizit auch die Neuerung „Einführung in Wirtschaft und Recht“ und „integrierte Informatik“ untersucht werden.

5.2 Die für EVAMAR II ausgewählten bzw. in Auftrag gegebenen Untersuchungen

Die Durchführung aller unter (2) skizzierten Untersuchungen war in der verfügbaren Untersuchungszeit aus verschiedenen (praktischen, nicht grundsätzlichen) Gründen nicht machbar. Deshalb musste für EVAMAR II eine Auswahl getroffen werden. Gemessen am umfassenden Modell handelt es sich um eine Minimalvariante, die aber noch finanzierbar war. Das Ergebnis der Auswahl ist in Abbildung I.2 dargestellt.

Abb. I.2: Für EVAMAR II ausgewählte Untersuchungen



Legende:
- Die Buchstaben A-D entsprechen den Teilprojekten.
- UF: Abkürzung für „Unterrichtsfach“
- TP: Abkürzung für „Teilprojekt“

Im Rahmen von EVAMAR II sollte die Konkretisierung der Ziele der Maturitätsbildung für ausgewählte Elemente des Hauptzieles der Studierfähigkeit erfolgen. Das zweite Hauptziel (vertiefte Gesellschaftsreife durch breite Allgemeinbildung) wurde damit nur teilweise in die durch den Auftraggeber gewählte Ziel-Operationalisierung einbezogen (im Überschneidungsbereich) und sollte in weiteren Projekten (ausserhalb von EVAMAR II) bearbeitet werden.

Zur Operationalisierung des Konstrukts der Studierfähigkeit waren grundsätzlich sowohl die fachübergreifenden wie auch die fachspezifischen Kenntnisse, Fähigkeiten, Fertigkeiten und Bereitschaften für ein erfolgreiches Studium zu ermitteln. Im Zentrum stand dabei die Frage, was von den Studierenden an tatsächlichen Lern- und Denkprozessen beim Lernen und Arbeiten an der Hochschule gefordert wird und über welche fachlichen und überfachlichen Kompetenzen sie dabei verfügen müssen. Dazu wurden für EVAMAR II aus dem in Abschnitt 2 beschriebenen umfassenden Konzept die folgenden Untersuchungen ausgewählt:

A. Untersuchungen zur Klärung des Konstrukts „Hochschulreife“ bzw. „Studierfähigkeit“ aus Sicht der Universitäten.

   A1. Inhaltsanalyse der Fachcurricula, Skripten, zentralen Lehrbücher usw. an Universitäten: Ermittlung der vorausgesetzten fachlichen Anforderungen verschiedener Studienfächer.

   A2. Analyse von relevanten Aufnahmeprüfungen: Ermittlung der kognitiven Anforderungen.

   A3. Inhaltsanalyse von ersten Prüfungen an den Universitäten (Zwischenprüfungen und Assessment-Stufen-Prüfungen): Ermittlung der kognitiven Anforderungen.

   A4. Ergänzende Befragung von Dozierenden von Grundsemestervorlesungen über Studienanforderungen und -voraussetzungen.

B. Operationalisiertes Ziel „Studierfähigkeit“ der Maturitätsbildung: Entwicklung von Kompetenzrastern zur Systematisierung der Ergebnisse der vorangehenden Untersuchungen.

Aus den Untersuchungen A und B sollten somit wesentliche Komponenten faktischer Anforderungen an die Studierenden bei Studienbeginn resultieren. Zu beachten ist bei diesen Untersuchungen zur Bestimmung des Evaluationskriteriums „Studierfähigkeit“, dass die Ergebnisse nicht zwingend einem auch beabsichtigten bestimmten Optimum entsprechen müssen, weil die ermittelten Anforderungen bereits die Anpassung an nicht optimale Voraussetzungen der Studierenden beinhalten können. Zudem befinden sich zurzeit die Anforderungen der Hochschulen selbst in einem dynamischen Prozess (Bologna-Reform). Die faktischen Anforderungen der universitären Hochschulen an die Studierenden sollten deshalb auch nach EVAMAR II im Rahmen eines längerfristigen Monitorings periodisch neu ermittelt (Top Down) und einem kritischen Diskurs zugeführt werden (Bottom Up). In diesem Zusammenhang könnte dann auch eine allfällige Diskussion über Bildungsstandards am Gymnasium erfolgen.

Für die Untersuchung der Erreichung der Zielgrösse „Studierfähigkeit“ wurden für EVAMAR II aus dem umfassenden Gesamtkonzept (siehe Abschnitt 5.1) die folgenden weiteren Teilprojekte ausgewählt:

C. Ergebnisse der Maturitätsbildung: Im Zentrum von Teilprojekt C, welches an die Ergebnisse der Teilprojekte A und B anknüpft, stand die Entwicklung und Durchführung von Tests, welche in den drei Fächern Erstsprache, Mathematik und Biologie hochschulrelevante Teile der Ergebnisse der Maturitätsbildung messen. Diese wurden aufgrund von weiter hinten nachzulesenden Überlegungen (siehe Kapitel IV) ergänzt durch eine Zusatzbefragung und – an einer Subgruppe von Gymnasien – durch einen Testteil, der hochschulrelevante, allgemeine kognitive Fähigkeiten erfasst. In die Auswertungen wurden zudem die effektiv erzielten Maturitätsnoten einbezogen.

   Damit beschränkte sich im Vergleich zum umfassenden Konzept die Anzahl der Fächer, für die Tests zur Leistungsmessung erstellt und durchgeführt werden sollten. Im Sinne einer Minimalvariante waren es Mathematik, Erstsprache und ein naturwissenschaftliches Fach. Erstsprache und Mathematik werden in der einschlägigen Literatur – zusammen mit Englisch – als jene Fächer bezeichnet, in denen u. a. basale Fähig-

keiten erworben werden, die für alle Studienfächer von Bedeutung sind (z. B. Huber, 1997 und 1998). Dass nicht wie vom Auftragnehmer vorgeschlagen Englisch, sondern ein naturwissenschaftliches Fach einzubeziehen war, wurde durch den Auftraggeber entschieden. Die Naturwissenschaften stehen nämlich direkt im Zentrum der bildungspolitischen Diskussion um das MAR 95 (1995). Die vorgenommene Fächerauswahl entspricht somit auch einem politischen Konsens der Auftraggeber. Als naturwissenschaftliches Fach wurde später Biologie ausgewählt (zur Begründung dieser Wahl siehe Kapitel IV).

D. Qualität der Maturitätsprüfungen und der Maturaarbeit: Mit Teilprojekt D sollten Aspekte der an den Gymnasien mittels der Maturitätsprüfungen schulintern vorgenommenen Leistungsmessungen untersucht werden. Dabei ist zu beachten, dass die Maturitätsnoten die Indikatoren des erst letzten Gliedes „Maturaprüfungen" in der Kette der maturitären Qualifizierung sind und nur eine beschränkte Aussagekraft besitzen. Sie geben nicht notwendigerweise die effektiven Kompetenzen der Maturandinnen und Maturanden wieder. Zwischen den Zielvorgaben von Art. 5 MAR sowie dem RLP einerseits und den aus den Noten-Ergebnissen der Maturitätsprüfungen interpretierbaren Qualitäten andererseits sind bekanntlich viele Einflussgrössen wirksam, die in entsprechenden Untersuchungen erfasst und zueinander in Beziehung gesetzt werden könnten. Es sind dies z. B. (1) die Umsetzung der Vorgaben des MAR 95 (1995) in den lokalen Stundentafeln und Lehrplänen, (2) die Qualität des Unterrichts und (3) die Qualität der Maturitätsprüfungen und der Maturaarbeiten. Die Untersuchungen von EVAMAR II beschränkten sich aus Ressourcengründen auf den dritten Aspekt. Somit wurden die folgenden Teilprojekte durchgeführt:

   D1. Inhaltsanalyse von Maturaprüfungen (fachinhaltlich und bezüglich kognitivem Anspruchsniveau) und Vergleich der Ergebnisse mit den in Teilprojekt B ermittelten Zielkompetenzen unter Einbezug der Ergebnisse der zentralen Leistungsmessung gemäss Teilprojekt C.

   D2. Analyse von Maturaarbeiten: Untersuchung der Frage, inwieweit Maturandinnen und Maturanden mittels der Maturitätsarbeiten gelernt haben, wissenschaftlich zu arbeiten; dies geschah durch Nachbewertung der Arbeiten (Expertenrating), Befragung der verfassenden Maturandinnen und Maturanden sowie Rekonstruktion institutioneller Rahmenbedingungen.

5.3 Grenzen der Studie EVAMAR II

Aus dem Vergleich des umfassenden Konzepts (Abschnitt 5.1) mit den daraus ausgewählten Teilen (Abschnitt 5.2) ergibt sich, dass es sich bei EVAMAR II nicht um eine vollständige Evaluation des Gymnasiums handelt. Grundsätzlich sind nur Aussagen über einige ausgewählte Aspekte der Ziele und der Zielerreichung der gymnasialen Bildung möglich sowie Erkenntnisse über ausgesuchte Gesichtspunkte der Schnittstelle Gymnasium-Universität.

Zunächst handelt es sich nur um Wissen und Können, welches auch mit Lernprodukt- und Fähigkeitstests sowie mittels Selbsteinschätzungen erfasst werden kann. Wissen und Können im Sinne eines umfassenden Kompetenzbegriffs sensu Weinert (2001) lässt sich aus methodischen und ressourcenbedingten Gründen im Rahmen des Projekts nicht messen. Dazu gehören zum Beispiel Fähigkeiten im Bereich der Erkenntnisgewinnung (Hypothetisieren, praktische Arbeitstechniken wie Experimentieren, anschliessende Dateninterpretation etc.), schriftliche und mündliche Kommunikationsfähigkeit, Fähigkeiten der kriterienbasierten Beurteilung komplexer Sachverhalte sowie allgemein schöpferisch-kreative Fähigkeiten. In diesem Zusammenhang kann aber darauf hingewiesen werden, dass aus kognitionspsychologischer Sicht der Verfügbarkeit spezifischen Fachwissens als Grundlage des Verstehens, des weiteren Lernens wie auch des raschen Erinnerns eine viel grössere Bedeutung zukommt, als dies zuweilen in der umfassenden Diskussion über Kompetenzen suggeriert wird. Mittlerweile sind die Vorteile eines soliden Fachwissens für das Verstehen, Lernen und Erinnern gerade im Zusammenhang mit Fachtexten empirisch gut belegt (z. B. Weinert, 2001; siehe auch die ausführliche Darstellung in Kapitel IV, Abschnitt 1.2.1). Die im Gymnasium erlangte Qualität des Fachwissens ist im Hinblick auf die pragmatische, aber unter keinen Umständen zu vernachlässigende Seite der „allgemeinen Studierfähigkeit" besonders relevant

für EVAMAR II, zumal dieses Projekt den Fokus auf die Schnittstelle Gymnasium-Universität richtet, also besonders auf die Studierfähigkeit ganz zu Beginn des Studiums. Die vordergründig „banalen“ Fähigkeiten des Verstehens, Lernens und Erinnerns von Fachtexten dürften für das erfolgreiche Bestehen zumindest der allerersten universitären Prüfungshürden eher von ausschlaggebender Bedeutung sein als die Qualität der mehr handlungs- (und berufsorientierten) Fähigkeiten der Erkenntnisgewinnung, Kommunikationsfähigkeit und Beurteilung.

Aufgrund der methodischen und ressourcenbedingten Grenzen des Projekts handelt es sich im Weiteren nicht um das gesamte an den Universitäten vorausgesetzte Fachwissen und -können, sondern nur um das aufgrund der zugänglichen Lehr- und Prüfungsmaterialien und der Einschätzung der im Projekt mitarbeitenden Codier-Experten erfassbare bzw. erfasste. Im Hinblick auf die Verwendung der Ergebnisse aus der Materialanalyse (Teilprojekt A) im Teilprojekt C (Entwicklung und Durchführung von Tests) war methodisch selbst aus dieser beschränkten Teilmenge noch eine Auswahl zu treffen, da die Tests sonst viel zu umfangreich ausgefallen wären; es handelt sich dennoch um eine beachtliche Auswahl des effektiv für die Aufnahme eines beliebigen Studiums notwendigen Eingangswissens und -könnens.

Folgende Komponenten einer umfassenden Studier- und einer vertieften Gesellschaftsfähigkeit werden also im Projekt EVAMAR II nicht untersucht und nicht auf ihr Vorhandensein bei Maturandinnen und Maturanden analysiert:

- Wissen und Können, welches zwar Teil einer breiten Allgemeinbildung ist – als Grundlage zur Lösung anspruchsvoller Aufgaben in der Gesellschaft (vertiefte Gesellschaftsreife) – und damit in der gymnasialen Bildung vermittelt werden muss, aber an der Universität nicht direkt vorausgesetzt wird,
- weiteres notwendiges Eingangswissen für diejenigen Studienfächer, welche nicht zu den ausgewählten gehören (siehe Kapitel II, Abschnitt 1),
- zur Studierfähigkeit gehörendes notwendiges Wissen und Können aus anderen Unterrichtsfächern als Erstsprache, Mathematik und Biologie,
- Elemente überfachlicher Kompetenzen, die im gegebenen Projektrahmen nur schlecht erfassbar gewesen wären, dennoch aber für die erfolgreiche Bewältigung eines Studiums wichtig sind (z. B. die Fähigkeit des selbst organisierten Lernens, der allgemeinen Selbstorganisation, des Recherchierens, der raschen Verarbeitung umfangreicher Fachliteratur usw.). Die Bedeutung solcher Faktoren liesse sich nur durch weitere, im umfassenden Konzept vorgeschlagene Untersuchungen ermitteln (Längsschnittuntersuchungen bei Studierenden), und sie wären nur durch aufwändige Assessmentverfahren zu messen. Immerhin wurden solche Faktoren aber bereits im Projekt EVAMAR I (siehe Abschnitt 2) und in weiteren Evaluationsstudien (siehe Abschnitt 3) über Befragungen, also über Selbsteinschätzungen und damit annäherungsweise valide ermittelt. Im Zusatzfragebogen von EVAMAR II wurden zudem diesbezüglich ergänzende Fragen zu EVAMAR I eingestreut.

Auf der Grundlage der für EVAMAR II ausgewählten Untersuchungen können somit nur über *Teil*aspekte der Studierfähigkeit Aussagen gemacht werden. Das Ziel, „auf anspruchsvolle Aufgaben in der Gesellschaft vorzubereiten“, wurde zudem nicht direkt untersucht. Diese Einschränkungen sind bei der Interpretation der Ergebnisse und bei der Ableitung von Massnahmen für die künftige Gestaltung des Gymnasiums zu beachten!

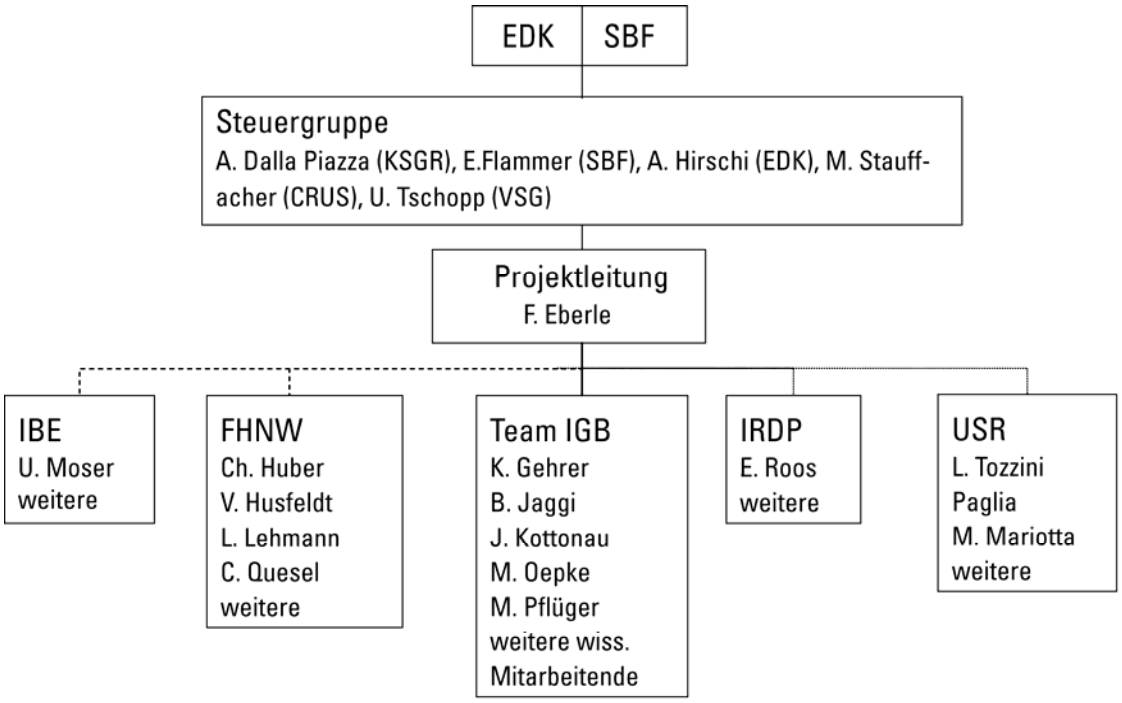
In der Diskussion um das Projekt EVAMAR II war häufig vernehmbar, dass in den Teilprojekten A und B die künftigen Bildungsstandards für das Gymnasium festgelegt würden. Das scheint ein Missverständnis zu sein. In den Teilprojekten A und B handelt es sich um eine Untersuchung des Ist-Zustandes eines Teiles der Schnittstelle Gymnasium-Universität. Den Ergebnissen kann keine Soll-Funktion zukommen. Zur Festlegung der Soll-Anforderungen wären zusätzliche und möglichst breit abgestützte Abstimmungsprozesse zwischen den an der Schnittstelle beteiligten Institutionen notwendig, die erst dann in von Universitäten und Gymnasien gemeinsam getragene Bildungsstandards münden könnten. Erst dann würde es sich um offizielle, von den zuständigen Gremien neu überdachte und verbindlich beschlossene künftige Leistungsanforderungen handeln. Bei ihrer Festlegung müsste man ausserdem die ganze Breite der in Art. 5 des MAR 95 (1995) vorgegebenen Ziele gymnasialer

Bildung diskutieren. Die Ergebnisse von EVAMAR II können somit für allfällige Arbeiten zur Entwicklung von Bildungsstandards nur teilweise als Grundlage dienen.

5.4 Projektorganisation und Dank

Die Aufbauorganisation des Projekts ist aus dem Organigramm in Abbildung I.3 ersichtlich.

Abb. I.3: Organigramm von EVAMAR II



Legende:
- EDK: Schweizerische Konferenz der kantonalen Erziehungsdirektoren
- SBF: Staatssekretariat für Bildung und Forschung
- KSGR: Konferenz Schweizerischer Gymnasialrektorinnen und Gymnasialrektoren
- CRUS: Rektorenkonferenz der Schweizer Universitäten
- VSG: Verein Schweizerischer Gymnasiallehrerinnen und Gymnasiallehrer
- IBE: Institut für Bildungsevaluation an der Universität Zürich
- IGB: Institut für Gymnasial- und Berufspädagogik der Universität Zürich
- FHNW: Fachhochschule Nordwestschweiz, Pädagogische Hochschule, Institut Forschung und Entwicklung
- IRDP: Institut de recherche et de documentation pédagogique, Neuchâtel
- USR: Ufficio studi e ricerche, Bellinzona

Der Auftrag zur Projektleitung von EVAMAR II ging an Prof. Franz Eberle vom Institut für Gymnasial- und Berufspädagogik (IGB, vormals Höheres Lehramt Mittelschulen) der Universität Zürich.

Dem Kernteam des IGB gehörten folgende wissenschaftliche Mitarbeiterinnen und Mitarbeiter an: Nina Barske (bis November 2007), Karin Gehrer (Erstsprache), Daniel Halter (bis Juni 2006), Beat Jaggi (von August 2006 bis Juli 2008, Mathematik), Johannes Kottonau (ab Juli 2006, Biologie), Maren Oepke (operative Leitung, Erstsprache) und Michael Pflüger. Die Angaben in Klammer bei einzelnen Personen weisen auf die spezifischen fachlichen und organisatorischen Zuständigkeiten hin. Für andere spezifische Aufgaben waren viele weitere wissenschaftliche Mitarbeitende temporär in das Projekt eingebunden. Namentlich erwähnt seien Dirk Bauer, Christine van Bedaf und Claudia Illés. Ruth Berri vom Lehrstuhlsekretariat hat grosse Teile der administrativen Arbeiten ausgeführt.

Für die Romandie haben die Kooperationspartner des Institut de recherche et de documentation pédagogique (IRDP) in Neuchâtel, namentlich Matthis Behrens als Direktor und Eva Roos als wissenschaftliche Mitarbeiterin, Aufgaben übernommen, uns insbesondere bei der Entwicklung und Anpassung der Instrumente unterstützt, die Durchführung der Datenerhebung in der Romandie sichergestellt und bei der Datenauswertung mitgearbeitet.

Im Tessin haben uns die Partner vom Ufficio studi e ricerche (USR) in Bellinzona, namentlich Emanuele Berger (Direktor), Luana Tozzini Paglia und Myrta Mariotta, unterstützt.

An den Tests des Teilprojekts C, insbesondere an der Stichprobenziehung, der Aufbereitung der Pretests und der Datenmodellierung nach dem Rasch-Modell, waren Urs Moser und Mitarbeitende vom Institut für Bildungsevaluation Zürich (IBE) massgeblich beteiligt.

Im Zusammenhang mit dem überfachlichen Fähigkeitstest (UFT) erfolgte eine Zusammenarbeit mit Prof. Klaus-Dieter Hänsgen vom Zentrum für Testentwicklung und Diagnostik Fribourg (ZTD).

Das Teilprojekt „Analyse der Maturaarbeiten – TP D2“ wurde vollständig von der Pädagogischen Hochschule der Fachhochschule Nordwestschweiz konzipiert und durchgeführt (Christina Huber, Prof. Vera Husfeldt, Lukas Lehmann und Prof. Carsten Quesel [Leitung]). Unterstützt wurde dieses Team von Felix Boller (PH FHNW) und Eva Roos (IRDP).

All den aufgeführten Personen möchten wir unseren grossen Dank aussprechen. Der EDK und dem SBF als Auftraggeber danken wir speziell für das Vertrauen und allen Mitgliedern der Steuergruppe für die äusserst konstruktive Begleitung des Projekts. Ein mindestens so grosser Dank gebührt all den Maturandinnen und Maturanden, die bereit waren, die Tests zu absolvieren und die Fragebögen auszufüllen oder die uns ihre Maturaarbeiten für das Rating zur Verfügung gestellt haben. Wir danken im Weiteren den zahlreichen Lehrpersonen, Mitgliedern von Schulleitungen und administrativen Mitarbeitenden, die uns bei der Erhebung der Daten aktiv unterstützt haben. Danken möchten wir schliesslich der Konferenz Schweizerischer Gymnasialrektorinnen und -rektoren (KSGR) sowie dem Verein Schweizerischer Gymnasiallehrerinnen und -lehrer (VSG), die uns mental unterstützt haben.

II Projektteil A: Elemente der Studierfähigkeit aus Sicht der Hochschulen

In den Teilprojekten A1 bis A4 sollte in einem ersten Hauptschritt der Evaluation – Festlegung der Evaluationskriterien bzw. der Zielgrössen der Maturitätsbildung – jenes Wissen und Können ermittelt werden, welches beim Eintritt in eine Schweizer Universität vorausgesetzt wird. Erst danach liess sich – in einem zweiten Hauptschritt des Evaluationsverfahrens – die Erreichung der Zielgrössen über die Ermittlung verschiedenster Indikatoren untersuchen. Die Zielgrössen leiteten sich aus den – nur in allgemeiner Form vorhandenen – gesetzlichen Zielen für Maturitätsschulen ab. So sind die Hauptziele der gymnasialen Bildung gemäss Art. 5 des MAR 95 (1995) die Erlangung „jener persönlichen Reife", die erstens „Voraussetzung für ein Hochschulstudium ist" (Hochschulreife oder Studierfähigkeit), und die zweitens „auf anspruchsvolle Aufgaben in der Gesellschaft vorbereitet" (vertiefte Gesellschaftsreife durch breite Allgemeinbildung). Diese Ziele sind nur teilweise deckungsgleich. Im Rahmen des Projekts EVAMAR II sollte die Konkretisierung der Ziele der Maturitätsbildung gemäss Auftrag nur für ausgewählte Aspekte des Hauptziels der Studierfähigkeit erfolgen. Das zweite Hauptziel (vertiefte Gesellschaftsreife durch breite Allgemeinbildung) wird damit nur im Überschneidungsbereich der beiden Ziele in die Ziel-Operationalisierung einbezogen (siehe im Einzelnen in Kapitel I, Abschnitt 5). Dabei kommen in den Teilprojekten unterschiedliche Methoden zum Einsatz. Im Folgenden werden die Datenbasis und die einzelnen Teilprojekte detailliert dargestellt.

1 Datenbasis der Teilprojekte TP A1, TP A3 und TP A4

Aus Ressourcengründen konnte keine Vollerhebung durchgeführt werden. Die Untersuchung erfasste die 16 meistgewählten Fächer (gemessen an der Zahl der Studierenden), die im Studienjahr 2004/2005 von etwa 70% aller Immatrikulierten in der Schweiz studiert wurden. Als Datengrundlage für diese Auswahl diente die Erhebung des Bundesamts für Statistik der Studierenden des Studienjahrs 2004/2005 (BfS, 2008a). Es stellte sich im Weiteren die Frage, ob es – wieder im Verhältnis zum Aufwand – überhaupt sinnvoll sei, vergleichbare Studiengänge mehrfach zu analysieren, da ähnliche Resultate wahrscheinlich waren. Deshalb konzentrierte sich die Auswahl der zu untersuchenden Studienfächer auf jene zwei oder drei Universitäten, an denen zusammen mehr als 50% aller Studierenden dieses Faches eingeschrieben waren. Aufgrund dieser zweistufigen Auswahl wurde insgesamt der Einbezug von Studiengängen geplant, in denen ca. 37% aller Studierenden in der Schweiz immatrikuliert sind. In Tabelle II.1 sind die ausgewählten Studienfächer aufgeführt, geordnet nach der Grösse der Studierendenzahlen, sowie die nach dem genannten Verfahren ausgewählten und einbezogenen Universitäten. Diese Auswahl bildete die Datenbasis für die Untersuchungen in den Teilprojekten TP A1, TP A3 sowie TP A4.

Tab. II.1: ausgewählte Studienfächer der Datenbasis von TP A

Fachrichtung	Universitäten
Wirtschaftswissenschaften	Universität ZH (UZH), Universität St. Gallen
Rechtswisssenschaften	UZH, Universität Freiburg
Psychologie	UZH, Universität Bern
Humanmedizin	UZH, Universität Lausanne
Geschichte	UZH, Universität Bern
Biologie	UZH, ETHZ, Universität Basel
Politikwissenschaften	Universität Genf, UZH
Architektur & Planung	ETHZ, Universität Lugano
Informatik	ETHZ, EPFL
Soziologie	Universität Freiburg, UZH
Germanistik	UZH, Universität Basel
Kommunikation & Medien	UZH, Universität Lugano
Französisch	Universität Lausanne, Universität Genf
Erziehungswissenschaften	Universität Genf, UZH
Englisch	UZH, Universität Bern
Physik	ETHZ, EPFL

2 Teilprojekte A1 und A3: Inhaltsanalyse von Studienmaterialien und Prüfungen

In TP A1 wurden Fachcurricula, Skripte sowie zentrale Lehrbücher der verschiedenen Studienfächer des ersten und zweiten Semesters einer Dokumentenanalyse unterzogen. Dabei ging es um die Ermittlung von Wissen und Können, das zum Verständnis und zur Bearbeitung der Lehrmaterialien vorausgesetzt wird. Dazu wurden ein Analyseraster und ein dazugehörendes Codier-Manual entwickelt, welches den untersuchenden Experten als Anleitung bei der Codierung der Materialien diente. Hinsichtlich der Analysekriterien musste nach solchen gesucht werden, die für die Untersuchung der sehr unterschiedlichen Studienmaterialien aller Studienfächer gleichermassen anwendbar schienen. Das fächerunspezifische Wissenskonzept von Bloom et al. (1956 und 1974) in der im Sinne einer Anpassung an neuere Erkenntnisse der Kognitionspsychologie sorgfältig überarbeiteten und erweiterten Fassung von Anderson et al. (2001) erwies sich als Basis für die Kategorienbildung als geeignet. Es ermöglichte vergleichende Aussagen über alle Fächer, und es handelt sich um eines der bestbewährten und meist verbreiteten pädagogisch-didaktischen Modelle der letzten Jahrzehnte[2] (siehe auch die Ausführungen dazu in Kapitel V, Abschnitt 1.2).

Da die Lehrmaterialien häufig auch Übungen und Aufgaben umfassen, die den Lernenden über die reine Wissensrezeption hinausgehende kognitive Fähigkeiten bei der Bearbeitung abverlangen, wurde für die Codierung von Aufgaben ein von der Codierung der Lehrtexte ausgehendes angepasstes Kategoriensystem entwickelt, das diese kognitiven Fähigkeiten integriert. Es zeigte sich, dass die Codierung der Aufgaben schliesslich für die in Teilprojekt A3 zu leistende Analyse der ersten Prüfungen im Studium übernommen werden konnte. Die nachfolgende Beschreibung der Methode bezieht sich daher auf beide Teilprojekte, A1 und A3, da ihnen auch die gleiche Datenbasis zugrunde liegt.

2.1 Methode und Instrumente TP A1 und TP A3

Die in EVAMAR II angewandte Methodik in den Projekten A (und B) ist neu, es konnte nicht auf ein bereits bestehendes Instrumentarium zurückgegriffen werden. Um die Materialien zu analysieren, wurden zwei Vorgehensweisen der Codierung entwickelt: das „feinmaschige“ und das „grobmaschige“ Codieren.

2.1.1 Die Methode des Feincodierens

Bei dieser Codierungsart wird unterschieden zwischen der Codierung von Lehrmaterialinhalten einerseits und den dazugehörenden Aufgaben bzw. Übungen und – wie bereits beschrieben – den Prüfungen (TP A3) andererseits.

Bei den Lehrmaterial-Inhalten wurden Sinneinheiten ermittelt, die Lernende bereits wissen müssen, um das Lehrmaterial zu verstehen. Sinneinheiten können sich aus bestimmten speziellen Wörtern (Nomen, Adjektiven, Adverbien etc.), Satzteilen, ganzen Sätzen und/oder mehreren Sätzen des Originaltextes ergeben und auch selbst in diesen Formen auftreten. Diese vom Lehrtext vorausgesetzten Sinneinheiten wurden in verschiedene Wissenskategorien eingeteilt, die sich in der ersten Dimension im Wesentlichen an den Kategorien von Bloom (1956 und 1974) bzw. Anderson et al. (2001) orientieren und in der zweiten Dimension an den Maturitätsfächern. Eine Sinneinheit kann dabei verschiedenen Schulfächern zugeordnet werden.

Tabelle II.2 stellt in vereinfachter Form das vorgegebene Raster dar, in welches die Sinneinheiten jeweils eingeordnet wurden. Während sich in den Zeilen die Wissenskategorien befinden (AA bis DB, zur Erläuterung siehe unten), sind in den Spalten die Abkürzungen für Schulfächer im Gymnasium aufgeführt, wobei nicht an allen

[2] Das ursprüngliche Modell wurde bislang in 22 Sprachen übersetzt.

Gymnasien alle Schulfächer gelehrt werden. Aus der Abbildung ist ersichtlich, dass neben den gängigen Wissensgebieten aus den Fachbereichen Deutsch (D), Französisch (F) , Italienisch (I), Englisch (E), Latein (L), Griechisch (G), Mathematik (M), Biologie (B), Chemie (Ch), Physik (Ph), Geschichte (G), Geographie (Gg), Wirtschaft & Recht (WR), Bildnerisches Gestalten (BG), Musik (Mu) sowie Philosophie, Pädagogik, Psychologie, Ethik und Religionslehre (PPP) zusätzlich als schulfachübergreifende Kategorien jene des Wissenschaftlichen Arbeitens (WA, z. B. Arten schriftlicher Arbeiten, Zeitplanung, Wissenschaftstheorie) sowie des Computer- und Informatikwissens (C) erfasst wurden. Die letzte Spalte „Zusätzliches" (Z) bot die Möglichkeit, solche Sinneinheiten zu erfassen, die sich nicht den anderen, schultypischen Kategorien zuordnen liessen (z. B. Symbole).

Tab. II.2: Vorgegebener Raster zur Erfassung der Wissensdimension und der Schulfächer

	D	F	I	E	L	Gr	M	B	Ch	Ph	G	Gg	WR	BG	Mu	PPP	WA	C	Z
AA																			
AB																			
BA		SE1																	
BB						SE2													
BC																			
CA																			
CB	SE3													SE3					
CC																			
DA																			
DB																			

Legende:
- SE = Sinneinheit
- Buchstaben in der Kopfzeile: Kürzel für die gymnasialen Unterrichtsfächer, wobei WA = Wissenschaftliches Arbeiten, C = Computer- und Informatikwissen, Z = Zusätzliches
- Buchstaben in der Kopfspalte: Kürzel für Wissenskategorien.

Neben der Zuordnung zu den in den Spalten aufgeführten gymnasialen Unterrichtsfächern wurden die Sinneinheiten dahin gehend beurteilt, in welche der in Tabelle II.3 dargestellten Wissenskategorien des Faktenwissens (AA, AB), des Konzeptionellen Wissens (BA bis BC), des Prozeduralen Wissens (CA bis CC) oder des Metakognitiven Wissens (DA und DB) sie sich entsprechend dem Kontext ihrer Textstelle einreihen liessen (Zeilen). Die ausführliche Beschreibung und die weitere Feineinteilung dieser bei der Codierung angewandten Wissenskategorien können im Codier-Manual nachgelesen werden (siehe technischer Bericht), wo auch Beispiele zu finden sind.

Tab. II.3: Wissenskategorien der Sinneinheiten (Hauptkategorien)

A. Faktenwissen
- AA Wissen über Terminologien
- AB. Wissen über allgemeine Details und Elemente

B. Konzeptionelles Wissen
- BA. Wissen über Klassifikationen und Kategorien
- BB. Wissen über Prinzipien und Generalisierungen
- BC. Wissen über Theorien, Modelle und Strukturen

C. Prozedurales Wissen
- CA. Wissen über (fach)spezifische „Skills"
- CB. Wissen von (fach)spezifischen Techniken und Methoden
- CC. Wissen von Kriterien für die Bestimmung einer angebrachten Prozedur

D. Metakognitives Wissen
- DA. Strategisches Wissen
- DB Wissen über kognitive Aufgaben, inkl. Kontextuellem Wissen und Bedingungswissen

Bei der Codierung der Aufgaben und Übungen sowie der Prüfungen (TP A3) wurde dieses Wissen ebenfalls erfasst, wenn dessen Kenntnis Voraussetzung für die Lösung und insofern Studien-Eingangswissen war, als zuvor keine Erklärung in den Lehrunterlagen erfolgt war. Dabei erfolgte die Zuordnung nur zu den Hauptwissenskategorien A bis D. Wurde in einer Aufgabe oder Prüfung hingegen nur auf Wissen Bezug genommen, das im Lehrmaterial vorher selbst vermittelt worden war, so erfolgte keine Erfassung auf der Wissensebene.

Neben der fallweisen Erfassung der Wissenskategorie wurde für jede Aufgabe und Übung bzw. Prüfung die für die Lösung der Aufgabe notwendigen kognitiven Fähigkeiten festgehalten. Die Kategorisierung dieser kognitiven Fähigkeiten orientiert sich ebenfalls an der Taxonomie von Bloom (1956 und 1974, weiterentwickelt von Anderson et al., 2001). Die Codierer erfassten die zur Lösung der Aufgaben vorzunehmenden Denkprozesse ebenso, wie dies bei der Analyse von Lehrinhalten geschah: mithilfe entsprechender Raster. Die Kategorisierung und Zuordnung kognitiver Fähigkeiten ist in Tabelle II.4 dargestellt. Eine ausführliche Beschreibung mit Beispielen kann wieder im Codier-Manual nachgelesen werden (siehe technischer Bericht).

Tab. II.4: Kategorien kognitiver Fähigkeiten für Aufgaben und Prüfungen

A.	Erinnern.
B.	Verstehen. Darunter fallen die folgenden Anforderungen: interpretieren (klären, paraphrasieren, übersetzen), beispielhaft erklären (illustrieren), klassifizieren (kategorisieren, subsumieren), zusammenfassen (abstrahieren, generalisieren), ableiten (schlussfolgern, vorhersagen, hochrechnen), vergleichen (kontrastieren), erklären (Modelle konstruieren).
C.	Anwenden. Dazu zählen: bearbeiten (ausführen), implementieren (anwenden), einfache Reflexion der eigenen Anwendung.
D.	Analysieren. In diese Kategorie fallen: differenzieren (unterscheiden, fokussieren, auswählen), organisieren (auffinden, aufzeigen von Zusammenhängen/Stimmigkeit, integrieren, skizzieren, strukturieren, bestimmen von Syntax/Grammatik), zuschreiben (Einzelteile bestimmen, zurückführen auf, zuordnen), Übertragung auf reale Situation (z. B. die reale Situation in die Sprache der Mathematik, Physik etc. übersetzen, dort entsprechend lösen, Ergebnis für die reale Situation interpretieren).
E.	Evaluieren bzw. Bewerten. Damit sind folgende Tätigkeiten gemeint: überprüfen (ermitteln, koordinieren, kontrollieren, überwachen, testen), kritisieren (beurteilen).
F.	Kreieren bzw. Erschaffen. Hierzu zählen: generieren (Aufstellen von Hypothesen, abstrahieren und kombinatorisch-schlussfolgernd denken), planen (entwerfen), produzieren (konstruieren), verfassen (im Sinne von: Beziehungen/Strukturen entdecken, erfinden).

Weiterhin wurde anhand bestimmter Kriterien eine Beschreibung und Beurteilung von Bildern und Grafiken vorgenommen sowie die Möglichkeit zur Codierung offener Kategorien geboten, sodass eine Erfassung möglich war, falls für das Verständnis von Texten oder das Lösen von Aufgaben bei den Studierenden weitere Kompetenzen, wie z. B. rhetorische oder motorische, erforderlich waren (vgl. dazu ausführlich das Manual im technischen Bericht).

Es handelt sich beim feinmaschigen Codieren insgesamt um ein sehr enges, detailgenaues Codierungsverfahren.

Das Codierungssystem wurde unter Mitwirkung der Codierer entwickelt und mehrfach diskutiert. Alle Codierer, die jeweils Experten für das von ihnen zu analysierende Studienfach sind, durchliefen eine entsprechende Schulung. Für die Reliabilitätsprüfungen wurden drei verschiedene Masse zur Feststellung der Inter-Rater-Reliabilitäten berechnet: die (paarweise erhobenen) Übereinstimmungen der Auswahl der Sinneinheiten, der Einordnung in die Wissenskategorie und der Zuordnung zu den Fächern. Es ergaben sich nur mittlere Werte ($r = .5$ bei der Auswahl der Sinneinheiten). Sie waren Anlass für Diskussionen und Nachschulungen zur Verbesserung der übereinstimmenden Codierung. Eine kontextbezogene Würdigung dieser Werte erfolgt weiter hinten (siehe Abschnitt 2.1.3).

2.1.2 Die Methode des grobmaschigen Codierens

Die zeitaufwändige feinmaschige Codierung wurde für eine intensive und genaue Detailerfassung der Einstiegskapitel in ein Lehrbuch oder Skript eingesetzt. Um ganze Modelle/Theorien und allgemein Sinneinheiten in einem grösseren Zusammenhang zu ermitteln, wurde für die nachfolgenden Kapitel desselben Lehrtextes ein grobmaschiges Verfahren entwickelt, das auch weniger zeitintensiv war. Während also die Codierung der ersten inhaltlichen Kapitel der Eingangslehrmaterialien und der Übungen mit dem bereits beschriebenen feinmaschigen Verfahren erfolgte, wurde das weitere Lehrmaterial einem schneller durchführbaren Analyse-Prozess unterzogen: dem „grobmaschigen Codieren“. Dabei ging es nach wie vor um die Suche nach Sinneinheiten, die Ler-

nende bereits wissen müssen, um das Lehrmaterial verstehen zu können (Eingangswissen). Beim grobmaschigen Codieren gehören aber nicht mehr nur bestimmte Wörter, Satzteile und/oder Sätze zu diesem Eingangswissen, sondern es kann sich auch um ganze Themen aus den zugrunde liegenden Schulfächern handeln (z. B. „Mathematik: Im Rahmen der Differenzialrechnung Funktionen ableiten"). Der Fokus liegt bei diesem Verfahren – im Gegensatz zur kleinschrittigen Methode – auf der Analyse solcher Sinneinheiten, die vorausgesetzt werden, um die zentrale Lehrbotschaft, die Kernaussagen von Kapiteln, verstehen zu können. Dabei wurde der Überlegung Rechnung getragen, dass man für das Verständnis einer Textaussage nicht immer den ganzen Text bzw. jeden einzelnen Begriff kennen muss. So können zum Teil ganze Textpassagen unverstanden bleiben, ohne dass das Verständnis für die zentrale Lehrbotschaft zwingend darunter leidet. Die nach dieser Festlegung zu kategorisierenden Wissensbestände wurden ebenfalls Schulfächern zugeordnet und entsprechend in Rastern festgehalten.

2.1.3 Zur messtechnischen Qualität

Weil es sich bei dieser Untersuchung um ein qualitatives Verfahren handelt, mussten des Weiteren spezifische Überlegungen zur messtechnischen Qualität angestellt werden. Es folgen zunächst jene zu Reliabilität und Validität:

- Zur Reliabilität: Die nur mittleren Werte der Reliabilität hätten eventuell erhöht werden können, wenn jeweils eine zweite oder gar dritte Nach-Codierung erfolgt wäre. Das war in diesem Projekt ressourcenbedingt nicht möglich. Die nur mittleren Übereinstimmungswerte resultieren aus der Subjektivität der Einschätzung durch die Codierer, was bei den untersuchten Texten als vorausgesetztes Wissen und Können gelten kann. Es ist insbesondere schwierig, die Sinneinheiten scharf zu trennen hinsichtlich des Kriteriums, ob es sich um selbstverständliches Allgemeinwissen oder um anspruchsvolleres (gymnasiales) Wissen handelt. Im Gegensatz zu Letzterem sollte Ersteres nicht erfasst werden, weil es zwar ebenfalls unabdingbar ist, aber im Hinblick auf eine erfolgreiche Studienbewältigung nicht zu den selektiven Wissensbeständen gehört und i.d.R. ausserhalb des Lernens am Gymnasium erworben wird. Orientiert man sich am Modell einer Wissenspyramide, befindet sich das selbstverständliche Allgemeinwissen im unteren und mittleren, das anspruchsvolle Wissen im oberen Bereich. Es stellt sich deshalb die Frage der projektzielbezogenen Auswirkungen der nur mittleren Intercoder-Reliabilität. Auch wenn die Codierer unterschiedlich viele Sinneinheiten im weniger anspruchsvollen Bereich erfassten, konnte eine hohe Übereinstimmung bei den schwierigeren, für die weitere Verwendung eher relevanten Sinneinheiten festgestellt werden. Aus dem gefundenen Universum von Sinneinheiten musste für das Erstellen der Tests (Projekt C) ohnehin eine Auswahl getroffen werden. Die dabei verfolgte Orientierung an den anspruchsvollen Sinneinheiten kann damit begründet werden, dass diese die eigentliche Hürde für die erfolgreiche Studienbewältigung ausmachen. Die Auswahl des Wissens und Könnens in den Projekten B und C bezieht sich somit auf die anspruchsvollen Sinneinheiten mit höherer Intercoder-Reliabilität bei deren Ermittlung.
- Zur Validität: Entsprechend den obigen Darlegungen wird auch die Validität der zu entwickelnden Tests durch die Methodik der Codierung nicht beeinträchtigt. Dabei ist zu beachten, dass die für die Tests auszuwählenden Sinneinheiten nur eine Teilmenge des für ein Universitätsstudium erforderlichen Wissens darstellen. Diese Auswahl enthält jedoch auch nichts, was für ein Studium nicht nachgewiesenermassen erforderlich ist oder zumindest damit zusammenhängt. Als Ergebnis der Feincodierung liegen allerdings nur relativ unzusammenhängende einzelne Sinneinheiten vor, die noch kein kohärentes Bild des für ein bestimmtes Studium notwendigen Wissens und Könnens ergeben. Dazu war die weitere Aufbereitung im jeweiligen Verwendungszusammenhang (Projekt C) notwendig. Entschärft wird das Problem des Vorliegens atomistischer Sinneinheiten, die teilweise aus dem übergeordneten Sinnkontext gerissen sind, durch das grobmaschige Verfahren, da dieses eher die Erfassung gesamtheitlicher Themen ermöglichte. Dennoch ist natürlich das Resultat der Analysen noch nicht gleichbedeutend mit „der" Studierfähigkeit.

Sicherlich erfolgt das Verstehen von Begriffen im rekonstruierenden Sinne häufig unmittelbar während des Lesevorgangs, also wiederum aufgrund der Fähigkeit, aus dem Sinnzusammenhang heraus bislang unbekannte oder unverstandene Begriffe zu verstehen, die jedoch gemäss unserer Auswertungslogik eindeutig Eingangswissen darstellen. Gegen diesen Einwand ist zu sagen, dass diese Begriffe trotzdem bei einer bestimmten Gelegenheit bereits einmal erlernt worden sein müssen. Falls keine Gedächtnisspur vorhanden ist, fehlt ein richtiges Verständnis. Natürlich lassen sich solche Wissenslücken im Einzelfall durch den Studierenden nachschlagen. Wenn er das übermässig tun muss, hat er aber erhebliche Schwierigkeiten beim Studium. Die Untersuchung basiert deshalb auf dem Bild des „idealen" Studierenden, der alle notwendigen Wissensvoraussetzungen bereits in das Studium mitbringt. Diese Argumentation wird wesentlich gestützt durch die bereits weiter oben erwähnte Erkenntnis, dass die Verfügbarkeit von spezifischem Fachwissen zu den wichtigsten Grundlagen weiteren Lernens zählt (z. B. Weinert, 2001) (siehe auch die ausführliche theoretische Unterfütterung im Zusammenhang mit der Fundierung der Tests in Kapitel IV, Abschnitt 1.2.1).

Schliesslich muss berücksichtigt werden, dass es auch mit dieser Methodik nicht möglich ist, alle wichtigen Facetten der Lesekompetenz genügend genau zu erfassen. Dazu gehören die folgenden Fähigkeiten:

- Die Kernaussagen eines Textes nicht nur zu verstehen, sondern auch als solche zu erkennen (und damit auch Wichtiges von Unwichtigem zu unterscheiden).
- Inhalte, die nicht explizit ausgeführt werden, aus dem Gesamttext zu schlussfolgern (z. B. die Intention des Autors).
- Dem Gesamtinhalt und der Argumentation eines Textes zu folgen, selbst wenn die Sprachlichkeit hohe formale und inhaltliche Anforderungen an die Lesenden stellt (z. B. neuartige Gedankengänge, inhaltliche Komplexität, ungewohnter Sprachstil, Häufung von Fremdwörtern, komplexe Grammatik).
- Konkurrierende Inhalte bzw. Argumente zu erkennen und zu unterscheiden.
- Die Bedeutung eines unbekannten Begriffs aus dem Kontext des Textes und/oder mithilfe guter Latein- und Griechischkenntnisse zu erschliessen.
- Einen fortschreitenden Text auch dann zu verstehen, wenn er viel Ausgangswissen „produziert", also fortlaufend neue Fachbegriffe einführt.
- Einen Text kritisch zu beurteilen (z. B. hinsichtlich seines Gehalts und seiner Relevanz für eine spezifische Fragestellung).
- Den kontextspezifischen Gehalt eines Begriffes zu erfassen bzw. isoliertes Faktenwissen übergeordneten Zusammenhängen zuzuordnen. Vor allem bei Detailinformationen bzw. Einzelbegriffen geht es in der Regel nicht so sehr um eine genaue Begriffskenntnis als vielmehr um die Herstellung des dazugehörigen Kontextes.
- Sich in anspruchsvolle Texte systematisch „einzudenken", sich das dazugehörige Begriffsinventar zu erarbeiten und die zentralen Argumentationslinien und theoretischen Standpunkte nachzuvollziehen (basale systematische Lektürekompetenzen).
- Sich Strategien zu erarbeiten, mit neuem Wissen sinnvoll umzugehen, da viele Lehrtexte – z. B. die obligatorischen Lehrbücher in der Soziologie oder die publizistischen Lehrtexte auf der internetgestützten Lernplattform – viel Ausgangswissen produzieren.

Dennoch konnte das in den ausgewählten Lehrtexten enthaltene begriffsspezifische Faktenwissen relativ gut erfasst werden. Häufig gelang es beim grobmaschigen Codieren auch, die zu den Begriffen gehörigen übergeordneten Themenbereiche zu ermitteln. Dass zur Bearbeitung von Lehrtexten die Verfügbarkeit dieses Begriffswissens zwar unabdingbar, aber häufig nicht ausreichend ist bzw. dass dazu auch die oben beschriebenen, weiterführenden Lesekompetenzen erforderlich sind, musste dann bei der Gestaltung der Kompetenzraster im Teilprojekt B und bei der Konstruktion der Aufgaben in Teilprojekt C berücksichtigt werden.

2.2 Datenrücklauf in TP A1 und TP A3

Die eingegangenen Unterlagen umfassen alle ausgewählten 16 Fächer, obwohl nicht alle kontaktierten Dozierenden ihre Studienunterlagen und Prüfungen bereitstellten (trotz z. T. mehrfachen Nachfragen). Insgesamt konnten von den ursprünglich geplanten 33 Studiengängen schliesslich Materialien von 26 Studienfächern untersucht werden. Das Volumen der analysierten Materialien entspricht dabei einem Anteil von Studierenden an der Gesamtstudierendenzahl der Schweiz – bezogen auf das Studienjahr 2004/2005 – von ca. 31% (geplante Zahl: 37%).

Hinsichtlich der Analyse der ersten Prüfungen im Rahmen von TP A3 ist einschränkend noch festzuhalten, dass nicht für jedes der Studienfächer Prüfungen zur Verfügung standen, weil in manchen Studiengängen aufgrund der Studienstruktur innerhalb der analysierten ersten beiden Semester keine Prüfungen stattfanden.

2.3 Auswertungen in TP A1 und TP A3

Das Vorgehen bestand also in der Analyse der Unterrichtsmaterialien des ersten und zweiten Studiensemesters eines jeweiligen Studienfachs und im Dokumentieren der Sinn- bzw. Wissenseinheiten – zumeist pro Kapitel bzw. pro Aufgabe und Prüfung – in den beschriebenen Rastern, von denen auf diese Weise insgesamt über 1000 erstellt wurden. Die ermittelten Wissens- und Könnenselemente wurden auf der Grundlage der Codiertabellen in eine umfassende Datenbank eingelesen. Für die den einzelnen Maturitätsfächern zugeordneten Sinneinheiten ermöglicht die Datenbank Zusammenstellungen und Auswertungen sowohl auf der Ebene der einzelnen analysierten Lehrveranstaltungen als auch auf der Ebene des ganzen Studienfaches und ebenso auf der Ebene aller untersuchten Studienfächer.

Darüber hinaus entstand basierend auf den Rastern eine Verbindung von Studienfächern an den Universitäten und Schulfächern an den Gymnasien, die Einsicht in die Schnittstelle zwischen diesen beiden Institutionen innerhalb des schweizerischen Systems zulässt: So soll ja die Gesamtheit des in allen Schulfächern erworbenen Wissens für das Studium aller Studienfächer an Schweizer Universitäten qualifizieren. Über die Zusammenführung der Raster in der Datenbank konnten nun die beiden Ebenen der Studienfächer an der Universität und der Unterrichtsfächer an den Gymnasien (Maturitätsfächer) miteinander in Beziehung gesetzt werden. Diese Ebenen sind nicht symmetrisch und linear verknüpft, sondern komplex miteinander verwoben. Durch die matrixförmige Anordnung sollte der Gefahr der Vermischung dieser Ebenen oder der Gleichsetzung von Studien- mit Unterrichtsfächern (z. B. als nahtlose Fortsetzung einer fachspezifischen Ausbildung im entsprechenden Fach) entgangen werden. Ein einzelnes Unterrichtsfach dient vielmehr neben seiner allenfalls wissenschaftspropädeutischen Funktion für ein bestimmtes Studienfach[3] vor allem der Vorbereitung auf viele bzw. alle möglichen Studienfächer.

Nach der Sammlung des studienrelevanten Wissens und Könnens über die 16 Studienfächer hinweg war es nun möglich, die jeweils studienvorbereitenden Inhalte quer durch die Studienfächer unterrichtsfachweise zu bearbeiten und so in ein jeweils kohärentes Kompetenzraster für jedes Unterrichtsfach zu bringen. Als Resultat lässt sich für jedes Unterrichtsfach beschreiben, welchen Beitrag es für den Erwerb der Studierkompetenz – immer im Rahmen der effektiv untersuchten Materialien – leisten muss. Diese Beschreibung bildet nicht einfach den Rahmenlehrplan und die Fachlehrpläne an Maturitätsschulen ab. Die folgende Abbildung II.1 verdeutlicht den Zusammenhang von Schul- und Studienfächern, wie er sich als Ergebnis der Projekte A und B herstellen lässt. Die Studienfächer sind in den Zeilenköpfen und die Unterrichtsfächer in den Spaltenköpfen aufgeführt.

[3] Diese Funktion ist eigentlich gemäss Art. 5 des MAR 95 nicht vorgesehen (keine fachspezifische Vorbereitung), wird aber trotzdem von den Vertretern einiger Studienfächer angenommen.

Abb. II.1: Zusammenhang zwischen Studienfächern und Unterrichtsfächern

Wissen und Können aus →	Mathematik	Erstsprache	Geographie	Physik	Biologie	usw.: Bereich Y
für Studienfach ↓						
Germanistik	GM	GE	GG	GP	GB	GY
Recht	JM	JE	JG	JP	JB	JY
Soziologie	SM	SE	SG	SP	SB	SY
Biologie	BM	BE	BG	BP	BB	BY
usw.: X	XM	XE	XG	XP	XB	XY
	↓	↓	↓	↓	↓	↓
auszubildende Studierkompetenzen in	**Mathematik**	**Erstsprache**	Geographie	Physik	**Biologie**	Fach Y oder überfachlich
	↑	↑			↑	
elaborierte Kompetenzmodelle aus Mathematik, Erstsprache und Biologie						

Legende: GM = erforderliches Wissen und Können aus dem Maturitätsfach Mathematik (M) für das Studienfach Germanistik (G) usw.

2.4 Auswertungsergebnisse TP A1 und TP A3 und Interpretation

Wegen des Umfangs ist es an dieser Stelle nicht möglich, alles konkrete in den Codiertabellen und der Datenbank eingetragene Wissen und Können (unterrichtsfachweise) wiederzugeben. Insgesamt kann man über alle Studienfächer hinweg anhand qualitativer Analysen auf folgende Tendenzen hinsichtlich des vorausgesetzten Wissens und Könnens aus den verschiedenen Maturitätsfächern schliessen:

- Bei der Zuordnung zu verschiedenen Wissenskategorien überwiegen die Einträge in die Wissenskategorie Faktenwissen (AA) insgesamt deutlich, vorausgesetztes prozedurales oder metakognitives Wissen wurde dagegen eher selten gefunden. Diese Häufung gilt insbesondere für Eintragungen bei den sprachlichen Schulfächern und betrifft daher zunächst die drei Erstsprachen Deutsch, Französisch und Italienisch sowie auch Englisch. Viele Eintragungen sind aber auch in den Fächern Latein und Griechisch zu finden, wobei es sich zum grossen Teil um Doppeleintragungen zusätzlich zur Erstsprache handelt. Darin widerspiegelt sich der Zusammenhang der alten Sprachen mit der Begrifflichkeit in vielen Fachwissenschaften. Allerdings kann man diese Begrifflichkeiten auch verstehen bzw. kennen, wenn man die entsprechende alte Sprache im Gymnasium nicht belegt hat.
- Die Zuordnung der Einträge zu den einzelnen Schulfächern fällt sehr unterschiedlich aus und hängt erwartungsgemäss stark vom analysierten Studienfach ab. So häufen sich z. B. bei der Analyse des Studienfachs Biologie Einträge bei den Schulfächern Biologie und Chemie, oder die wirtschaftswissenschaftlichen Studiengänge bauen stark auf ein bereits überraschend umfangreich vorausgesetztes, wirtschaftliches Begriffsrepertoire auf. So ergeben sich beispielsweise für Studienanfänger mit dem Schwerpunktfach Biologie/Chemie oder Wirtschaft & Recht gewisse Startvorteile zu Beginn eines Studiums der Biologie bzw. der Wirtschaftswissenschaften.
- Griechisch- und Lateinkenntnisse können überwiegend als hilfreich, aber nicht unbedingt als notwendig angesehen werden, so etwa in den Studiengängen Rechtswissenschaften, Geschichte oder Pädagogik; eine profunde Kenntnis von Latein und Griechisch dürfte jedoch vor allem – wie bereits erwähnt – Vorteile für den Wortschatz insgesamt mit sich bringen, der sich wiederum über alle Studiengänge hinweg als sehr bedeutsam erweist. Dabei ist aber zu berücksichtigen, dass Fachtermini auch ohne ein entsprechendes Sprachenlernen verständlich sein bzw. gemacht werden können.
- Ob das die Wiedereinführung eines Obligatoriums für alte Sprache begründen kann oder der direkte Weg zur Aneignung des jeweils fachspezifischen Wortschatzes zu einer besseren Studierfähigkeit führt, muss an dieser Stelle offen bleiben.

- Neben den wichtigen Erstsprachkenntnissen und einem umfangreichen Fremdwortschatz wird von den Studierenden zum Teil ein gutes Leseverständnis gefordert: Insbesondere in den geistes- und sozialwissenschaftlichen Fächern Geschichte, Soziologie und Pädagogik häufen sich Eintragungen, die auf das Arbeiten mit komplexen Texten verweisen und eine entsprechende Lesekompetenz erfordern.
- Hinsichtlich der naturwissenschaftlichen Kenntnisse lassen sich bei allen nicht naturwissenschaftlichen Studienfächern (in unserem Falle alle Studienfächer ausser Physik, Biologie und Medizin) kaum Unterschiede in der Zahl der ermittelten Wissenselemente aus den Bereichen der Physik und der Chemie feststellen. Hingegen ist der Rückgriff auf biologische Begriffe und Konzepte in den Geistes- und Sozialwissenschaften deutlich häufiger als auf solche aus Physik und Chemie.
- Auch aus dem überfachlichen Bereich des Wissenschaftlichen Arbeitens werden Kenntnisse vorausgesetzt. Daneben werden aber auch in einigen Studienfächern Veranstaltungen angeboten, die in das Wissenschaftliche Arbeiten einführen.
- Die Analyse der Aufgaben und Prüfungen in Bezug auf das vorausgesetzte Wissen zeigt, dass meist auf wenig Vorwissen referiert wird, das bereits vom Gymnasium mitgebracht werden muss; als für die Prüfungen erforderliche kognitive Fähigkeiten wurden überwiegend die Fähigkeitsstufen „Erinnern von Wissen“, „Verstehen“, „Anwenden“ und „Analysieren“ befunden. Fähigkeiten des „Evaluierens, Kreierens“ oder des „Generierens und Planens“ werden hingegen von den Studierenden eher wenig gefordert, zumindest in den ersten beiden Studiensemestern.

Die in den Teilprojekten A1 und A3 gewonnenen Inhalte dienten als wichtigste Grundlage für die im weiteren Projektverlauf zu untersuchenden Schulfächer Erstsprache, Mathematik sowie Biologie (siehe Teilprojekte B und C bzw. Kapitel III und IV).

3 Teilprojekt A4: Dozierendenbefragung

3.1 Der Fragebogen

Gemäss dem Auftragskonzept für das Projekt (vgl. Eberle, 2005) waren insgesamt zehn leitfadengestützte Interviews bei je zwei Lehrstuhlinhabern oder Studienberatern aus der Philosophischen, Mathematisch-Naturwissenschaftlichen, Medizinischen, Wirtschaftswissenschaftlichen und der Rechtswissenschaftlichen Fakultät vorgesehen Im Verlaufe der Materialbeschaffung im Rahmen von Projekt A1 zeigte sich jedoch, dass es schwierig werden würde, geeignete Personen zu finden. Diese hätten dem folgenden Profil entsprechen müssen:

- Sie sollten den fachlichen Überblick über den Stoff aller Lehrveranstaltungen der ersten beiden Semester haben.
- Um Allgemeinplätze auf Fragen nach dem für das Studium notwendige Eingangswissen und -können zu vermeiden, sollten sich die Personen mit Didaktik auseinandergesetzt haben. Das bedeutet, dass sie bei der Planung von Lehrveranstaltungen etwa folgende Fragen klären könnten: Welche Voraussetzungen an Wissen und Können müssen Studierende in die einzelnen Veranstaltungen mitbringen? Wie lässt sich der Stand des Wissens und Könnens der Studierenden einschätzen?

Im Rahmen der Beschaffung der Unterlagen für die anderen Teilprojekte hat sich herauskristallisiert, dass wohl nur ganz wenige Personen über das für die Interviews notwendige Wissen für alle Lehrveranstaltungen des entsprechenden Studienfachs verfügen. Deshalb wurde in Absprache mit der Steuergruppe der Entscheid gefällt, alle Dozierenden der in den Teilprojekten A1 und A3 untersuchten Lehrveranstaltungen in die Befragung einzubeziehen. Eine Einschätzung des für die eigene Lehrveranstaltung vorausgesetzten Wissens und Könnens dürfte den meisten am ehesten gelingen. Weil sich damit die Anzahl der potenziellen Interviewpersonen auf über 300 erhöhte, musste die Befragung in schriftlicher Form vorgenommen werden. Gegenüber dem mündlichen Interview ergeben sich damit zwar methodische Nachteile, die Schriftlichkeit hat aber auch viele Vorzüge. Die gewichtigsten Vorteile liegen darin, die Befragung besser standardisieren und den befragten Dozierenden mehr Überlegungszeit gewähren zu können.

Ziel der Befragung war zu erfahren, welches fachliche und überfachliche Wissen und Können Dozierende bei Studierenden des ersten und zweiten Semesters für ihre jeweiligen Veranstaltungen voraussetzen. Darüber hinaus wurde erfragt, welche Lücken die Dozierenden beim fachlichen und überfachlichen Wissen der Studienanfängerinnen und -anfänger beobachten.

3.2 Datenbasis von TP A4 und Ablauf der Befragung

Die Befragung wurde an deutschschweizerischen Universitäten mit Dozierenden der Lehrveranstaltungen der ersten zwei Semester der in TP A1 und A3 bestimmten Studienfächer durchgeführt. Die Beschränkung auf die Deutschschweiz erfolgte in Absprache mit der Steuergruppe aus ressourcenökonomischen Überlegungen. Die Erhebung wurde durch eine Befragung ausgewählter bilingualer Unversitätsdozierender aus dem Tessin und der Romandie ergänzt, welche in der Folge mehrheitlich die Ergebnisse als übertragbar auf ihre Sprachregion einschätzten. Im Weiteren lässt sich post hoc festhalten, dass die Ergebnisse aus der Dozierendenbefragung die Testkonstruktion in Teilprojekt C nur unwesentlich beeinflusst haben, sie sollten hier lediglich einer allfälligen Grobkorrektur dienen, weil es für viele Dozierende schwierig sein dürfte, Eingangswissen und -können ihrer Lehrveranstaltungen präzise festzulegen. Eine solche Grobkorrektur war dann aber – mit Ausnahme der nachträglichen Aufnahme des Kompetenzbereichs „Sprachreflexion/Grammatische Kompetenz und Orthografie“ – nicht notwendig. Die Befragungsergebnisse sind jedoch als eigenständige Resultate im Hinblick auf Überlegungen zur Weiterentwicklung des Gymnasiums interessant, und sie finden als Interpretationshilfe in Teilprojekt C und in Teilprojekt D1 Verwendung. In späteren Analysen lassen sich eventuell noch Quervergleiche anstellen zwischen den Angaben, die die Dozierenden im Fragebogen machten und den Sinneinheiten, die sich aufgrund der Codierung von Materialien der Veranstaltungen dieser Dozierenden in der Datenbank befinden.

Insgesamt wurden 310 Fragebogen an Dozierende verschickt, die Veranstaltungen im ersten oder zweiten Semester durchführen, und die im Rahmen der in Tabelle II.1 genannten ausgewählten Studiengänge und an den für das TP A ausgewählten Universitäten tätig sind. Ausgefüllt wurden 140 Fragebogen, sodass die Rücklaufquote bei 45% liegt. Die Dozierenden verteilen sich wie folgt auf die Studienfächer: Germanistik (20), Anglistik (9), Geschichte (16), Psychologie (7), Pädagogik (12), Soziologie (3), Politikwissenschaften (4), Kommunikation & Medien (5), Wirtschaftswissenschaften (10), Rechtswissenschaften (16), Informatik (5), Physik (5), Biologie (16), Medizin (5), Architektur (7).

3.3 Ergebnisse der Befragung in TP A4

Es folgt eine Darstellung der wichtigsten Ergebnisse. Leere Antworten werden nicht extra ausgewiesen.

3.3.1 Bedeutung der gymnasialen Fachbereiche für die Lehrveranstaltungen (Frage 1)

Die Dozierenden wurden zunächst gebeten einzuschätzen, für wie wichtig sie fachspezifisches Wissen und Können aus den verschiedenen gymnasialen Maturafächern sowie in den Bereichen „Benutzerwissen Informatik“ und „Informatikfachwissen“ für das Verständnis und den erfolgreichen Besuch ihrer Veranstaltungen erachten. Es wurde ihnen folgende Frage gestellt:

„Aus welchen gymnasialen Fachbereichen müssen Studierende bereits Fachwissen und -können mitbringen, damit sie in der Lage sind, Ihrer Lehrveranstaltung zu folgen? Bitte beziehen Sie sich dabei auf Fachwissen und -können, das Studierende bereits beim Eintritt ins Studium mitbringen müssen, welches also nicht in anderen Lehrveranstaltungen im Studiengang erworben werden kann. Bitte kreuzen Sie pro Fachbereich ***das zutreffendste*** *Feld an.“*

Die Ergebnisse finden sich in Tabelle II.5. Zusätzlich zu den Prozentzahlen für die Antwortmöglichkeiten ist in der letzten Spalte jeweils der Mittelwert angegeben, gebildet aus der vierstufigen Skala. Die Zeilen sind nach diesem Wert sortiert.

Die Bedeutung der Vorbildung für die eigene Lehrveranstaltung wird in den verschiedenen Fächern recht unterschiedlich eingeschätzt. Ein Ranking ergibt die grösste Bedeutungszumessung für Deutsch, Englisch, Informatik-Benutzerwissen und Mathematik (in dieser Reihenfolge). Biologie und Physik befinden sich in dieser Reihung in der Mitte der Rangskala und Chemie am Beginn des letzten Drittels. Der Entscheid für den Einbezug von Deutsch bzw. Erstsprache und Mathematik in die Testfächerauswahl sowie Biologie als naturwissenschaftliches Fach wird somit durch die Befragung gestützt.

Tab. II.5: Bedeutung gymnasialer Fachbereiche für die eigene Lehrveranstaltung

gymnasiale Fächer oder Bereiche	kein Wissen und Können (1)	fragmenta-risches Wissen und Kön-nen (2)	Kenntnis wichtigster Grundlagen (3)	fundiertes Wis-sen und Können (4)	Mittel-wert
Deutsch (Erstsprache)	5.2%	6.0%	25.4%	63.4%	3.47
Englisch	16.7%	11.9%	43.7%	27.8%	2.83
Informatik-Benutzerw.	19.7%	18.9%	45.9%	15.6%	2.57
Mathematik	36.1%	15.6%	32.8%	15.6%	2.28
Geschichte	36.0%	22.4%	28.0%	13.6%	2.19
Philosophie	41.8%	30.3%	27.9%	.0%	1.86
Französisch	54.2%	15.8%	22.5%	7.5%	1.83
Wirtschaft & Recht	50.8%	26.7%	19.2%	3.3%	1.75
Geographie	56.0%	27.6%	14.7%	1.7%	1.62
Biologie	68.1%	11.8%	13.4%	6.7%	1.59
Physik	70.1%	10.3%	14.5%	5.1%	1.55
Ethik & Religion	60.8%	30.8%	8.3%	.0%	1.47
Latein	64.5%	28.1%	7.4%	.0%	1.43
Informatik-Fachw.	72.0%	18.6%	7.6%	1.7%	1.39
Psychologie	72.5%	18.3%	8.3%	.8%	1.38
Chemie	78.8%	12.7%	2.5%	5.9%	1.36
Italienisch	77.8%	13.7%	8.5%	.0%	1.31
Pädagogik	81.2%	12.0%	5.1%	1.7%	1.27
Bildn. Gestalten	83.8%	11.1%	4.3%	0.9%	1.22
Griechisch	89.5%	9.6%	.9%	.0%	1.11
Musik	91.4%	8.6%	.0%	.0%	1.09

N max. = 140

Schlüsselt man die Antworten auf die Frage nach den einzelnen Studienfächern auf, verändern sich die Positionen teilweise (siehe Tabelle II.6). Im Folgenden werden jeweils die ersten 4 Positionen sowie jene der EVAMAR-Testfächer (noch vor der Bestimmung des naturwissenschaftlichen Fachs) genannt.

- Für das Studienfach Germanistik werden Deutsch, Informatik-Benutzerwissen, Geschichte und Englisch als am wichtigsten bezeichnet. Biologie, Physik und Chemie rangieren an 10., 12. und 18. Stelle, Mathematik an 15.
- Bei Anglistik nehmen Englisch, Deutsch, Informatik-Benutzerwissen und Geschichte die ersten 4 Positionen ein. Biologie und Physik sind an 11. und Mathematik an 17. Stelle. Aus dem Fach Chemie wird kein Wissen und Können erwartet.
- Die Geschichts-Dozierenden werteten Deutsch, Geschichte, Englisch und Französisch als wichtigste Fächer für ein Geschichtsstudium. Mathematik und Biologie befinden sich auf den Plätzen 12 und 17. Wissen und Können aus Chemie und Physik werden nicht erwartet.

Tab. II.6: Bedeutung gymnasialer Fachbereiche nach Studienfächern (Rangzahl jeweils in Klammer)

gymnasiale Fächer oder Bereiche	Germanistik	Englisch	Geschichte	Psychologie	Pädagogik	Soziologie	Politikwissenschaften	Kommunikation & Medien	Wirtschaftswissenschaften	Rechtswissenschaften	Informatik	Physik	Biologie	Medizin	Architektur
Deutsch	3.60 (1)	2.78 (2)	3.88 (1)	3.83 (1)	3.45 (1)	2.67 (4)	3.75 (1)	3.75 (1)	3.50 (1)	3.93 (1)	2.80 (2)	2.25 (3)	3.47 (1)	3.75 (1)	2.67 (1)
Mathematik	1.23 (15)	1.14 (17)	1.44 (12)	3.17 (3)	2.00 (7)	3.33 (2)	2.75 (5)	3.00 (2)	2.64 (2)	1.85 (9)	3.60 (1)	3.40 (1)	3.13 (3)	2.75 (3)	2.20 (7)
Englisch	2.47 (4)	4.00 (1)	3.31 (3)	3.50 (2)	2.64 (3)	3.33 (2)	3.50 (2)	2.75 (4)	2.27 (5)	2.15 (7)	2.00 (4)	2.25 (3)	3.06 (4)	2.75 (3)	2.60 (2)
Biologie	1.38 (10)	1.29 (11)	1.06 (17)	2.00 (6)	1.40 (11)	1.00 (21)	1.25 (12)	1.00 (21)	1.10 (16)	1.08 (21)	1.20 (9)	1.25 (7)	3.24 (2)	3.25 (2)	1.00 (21)
Chemie	1.15 (18)	1.00 (21)	1.00 (21)	1.33 (11)	1.00 (21)	1.00 (21)	1.00 (21)	1.00 (21)	1.10 (16)	1.08 (21)	1.40 (6)	1.25 (7)	2.63 (5)	2.75 (3)	1.00 (21)
Physik	1.31 (12)	1.29 (11)	1.00 (21)	1.50 (10)	1.00 (21)	1.00 (21)	1.00 (21)	1.00 (21)	1.20 (12)	1.15 (14)	2.20 (3)	2.80 (2)	2.56 (6)	2.75 (3)	2.00 (8)
Informatik-Benutzerw.	2.50 (2)	2.63 (3)	2.75 (5)	2.50 (4)	3.00 (2)	3.67 (1)	3.00 (4)	3.00 (2)	2.40 (4)	2.43 (5)	1.40 (6)	2.00 (5)	2.56 (6)	2.33 (7)	2.60 (2)
Psychologie	1.36 (11)	1.29 (11)	1.06 (17)	2.17 (5)	2.09 (6)	1.33 (9)	1.00 (21)	1.67 (8)	1.30 (11)	1.38 (12)	1.00 (21)	1.00 (21)	1.06 (18)	2.00 (8)	1.50 (11)
Philosophie	2.19 (6)	1.67 (6)	2.31 (7)	1.67 (7)	2.00 (7)	1.67 (7)	2.75 (5)	1.00 (21)	1.60 (7)	2.31 (6)	1.20 (9)	1.25 (7)	1.25 (12)	1.67 (11)	2.25 (6)
Informatik-Fachw.	1.08 (21)	1.29 (11)	1.18 (15)	1.67 (7)	1.10 (14)	2.00 (5)	1.25 (12)	1.33 (9)	1.50 (8)	1.15 (14)	1.60 (5)	1.75 (6)	1.94 (8)	1.67 (11)	1.00 (21)
Pädagogik	1.15 (18)	1.14 (17)	1.19 (14)	1.60 (9)	2.45 (4)	1.33 (9)	1.00 (21)	1.00 (21)	1.20 (12)	1.08 (21)	1.00 (21)	1.00 (21)	1.06 (18)	1.33 (21)	1.25 (12)
Geschichte	2.50 (2)	2.11 (4)	3.35 (2)	1.33 (11)	2.45 (4)	2.00 (5)	3.25 (3)	2.25 (5)	1.70 (6)	2.62 (3)	1.00 (21)	1.25 (7)	1.13 (16)	2.00 (8)	2.50 (4)
Wirtschaft & Recht	1.23 (15)	1.14 (17)	2.25 (8)	1.17 (13)	1.60 (10)	1.67 (7)	2.25 (9)	2.25 (5)	2.45 (3)	2.50 (4)	1.40 (6)	1.25 (7)	1.19 (13)	1.33 (21)	1.75 (9)
Ethik & Religion	1.46 (9)	1.44 (9)	2.00 (9)	1.17 (13)	1.78 (9)	1.33 (9)	1.50 (10)	1.00 (21)	1.50 (8)	1.69 (10)	1.00 (21)	1.00 (21)	1.06 (18)	1.75 (10)	1.25 (12)
Bildn. Gestalten	1.08 (21)	1.57 (8)	1.00 (21)	1.00 (21)	1.00 (21)	1.00 (21)	1.00 (21)	1.00 (21)	1.20 (12)	1.08 (21)	1.20 (9)	1.00 (21)	1.50 (9)	1.67 (11)	2.40 (5)
Geographie	1.62 (8)	1.29 (11)	2.69 (6)	1.00 (21)	1.22 (13)	1.33 (9)	2.50 (7)	2.00 (7)	1.40 (10)	1.46 (11)	1.20 (9)	1.00 (21)	1.44 (10)	1.67 (11)	1.75 (9)
Französisch	2.21 (5)	1.63 (7)	2.88 (4)	1.00 (21)	1.30 (12)	1.33 (9)	2.50 (7)	1.00 (21)	1.10 (16)	3.00 (2)	1.00 (21)	1.00 (21)	1.44 (10)	1.33 (21)	1.25 (12)
Latein	1.93 (7)	1.75 (5)	2.00 (9)	1.00 (21)	1.00 (21)	1.33 (9)	1.50 (10)	1.00 (21)	1.20 (12)	1.31 (13)	1.20 (9)	1.00 (21)	1.19 (13)	1.67 (11)	1.00 (21)
Griechisch	1.23 (15)	1.00 (21)	1.20 (13)	1.00 (21)	1.00 (21)	1.00 (21)	1.25 (12)	1.00 (21)	1.10 (16)	1.08 (21)	1.20 (9)	1.00 (21)	1.07 (17)	1.67 (11)	1.00 (21)
Italienisch	1.31 (12)	1.38 (10)	1.73 (11)	1.00 (21)	1.10 (14)	1.33 (9)	1.00 (21)	1.00 (21)	1.00 (21)	1.92 (8)	1.00 (21)	1.00 (21)	1.19 (13)	1.33 (21)	1.00 (21)
Musik	1.25 (14)	1.29 (11)	1.13 (16)	1.00 (21)	1.00 (21)	1.00 (21)	1.25 (12)	1.00 (21)	1.00 (21)	1.08 (21)	1.00 (21)	1.00 (21)	1.00 (21)	1.33 (21)	1.00 (21)

- Gemäss den Angaben der Psychologie-Dozierenden muss für ein Studium in ihrem Fach am meisten Wissen und Können aus den Fächern Deutsch, Englisch und Mathematik mitgebracht werden. An 4. Stelle folgt Informatik-Benutzerwissen. Biologie steht auf Platz 6, Physik und Chemie an 10. und 11. Stelle.
- Als Grundlage eines Pädagogik-Studiums wurden Deutsch, Informatik-Benutzerwissen, Englisch, Pädagogik und Geschichte als am wichtigsten eingeschätzt. Mathematik und Biologie rangieren an 7. und 11. Stelle. Wissen und Können aus Chemie und Physik braucht es aus Sicht der Dozierenden nicht.
- Die Soziologie-Dozierenden setzten Informatik-Benutzerwissen, Mathematik, Englisch und Deutsch auf die vordersten Positionen. Physik, Chemie und Biologie haben keine Bedeutung.
- Für ein Studium im Fach Politikwissenschaften wurde von den entsprechenden Dozierenden Wissen und Können aus den Fächern Deutsch, Englisch und Geschichte sowie Informatik-Benutzerwissen als am wichtigsten bewertet. Mathematik ist an 5. Stelle zu finden, Biologie an 12. Position. Physik und Chemie braucht es nach ihrer Einschätzung nicht.
- Die Dozierenden des Studienfachs Kommunikation & Medien erwarten das umfassendste Wissen und Können aus den Fächern Deutsch, Mathematik und Englisch, sowie ein gutes Informatik-Benutzerwissen. Biologie, Chemie und Physik sind ohne Bedeutung.
- Die Dozierenden der Wirtschaftswissenschaften bezeichneten Wissen und Können aus den Fächern Deutsch, Mathematik und Wirtschaft & Recht sowie Informatik-Benutzerwissen als am wichtigsten für ihr Studium. Weniger Bedeutung hat das Fach Physik (Rang 12), und fast keine haben Chemie und Biologie (beide Rang 16).
- Bei den Rechtswissenschaften nehmen die Fächer Deutsch, Französisch, Geschichte sowie Wirtschaft & Recht die vorderen Plätze ein. Hingegen braucht es im Urteil der Dozierenden kaum Eingangswissen aus Physik (14), Chemie und Biologie (je 21).
- Für ein Studium der Informatik braucht es nach Einschätzung der Dozierenden vor allem Wissen und Können aus den Fächern Mathematik, Deutsch, Physik und Englisch. Chemie steht an 6., Biologie an 9. Stelle.
- In der Einschätzung der Physik-Dozierenden ist für ihr Studium Wissen und Können aus den Fächern Mathematik, Physik, Englisch und Deutsch am wichtigsten. Chemie und Biologie werden als gleich wichtig eingeschätzt (Position 7).
- Die Dozierenden des Studienfachs Biologie bezeichneten vor allem Wissen und Können aus den Fächern Deutsch, Biologie, Mathematik und Englisch als am wichtigsten. Chemie und Physik folgen auf den nächsten beiden Plätzen 5 und 6.
- Dozierende der medizinischen Fakultät bewerteten die Schulfächer Deutsch (Rang 1), Biologie (Rang 2), Englisch, Physik, Chemie und Mathematik (je Rang 3) als am wichtigsten für ihre Lehrveranstaltungen.
- Beim Studienfach Architektur schliesslich nehmen Deutsch, Informatik-Benutzerwissen, Englisch und Geschichte die ersten 4 Positionen ein. Mathematik befindet sich an 7. und Physik an 8. Stelle. Chemie und Biologie wurden als bedeutungslos eingeschätzt.

Insgesamt verbleibt somit das gymnasiale Fach Deutsch bzw. Erstsprache beim Vergleich des Mittelwerts der Bedeutungseinschätzung über alle Studienfächer hinweg überwiegend an erster Stelle und Englisch sowie Informatik-Benutzerwissen auf den vorderen Rängen. Bei den anderen Fächern ergeben sich teilweise erhebliche Unterschiede in der Rangierung. Mathematik wird in den sprach- und geisteswissenschaftlichen Fächern erwartungsgemäss als wesentlich weniger wichtig eingestuft, bleibt aber für viele Studienfächer sehr bedeutsam. An vorderer Stelle ist jeweils auch jenes Maturafach zu finden, das dem eigenen Studienfach entspricht.

3.3.2 Konkrete Wissensinhalte aus den für die Lehrveranstaltung bedeutsamen gymnasialen Fachbereichen (Frage 2)

Die Dozierenden wurden im Weiteren mittels folgender offener Frage nach konkretem Wissen und Können gefragt, auf welches ihre Lehrveranstaltung aufbaut:

„Bitte beschreiben Sie für jedes der in Frage 1 mit ‚Fundiertes Fachwissen und -können' oder ‚Kenntnis wichtigster Grundlagen' angekreuzten Fächer möglichst genau, welches Wissen und Können Sie in Ihrer Lehrveranstaltung voraussetzen. Bitte beschreiben Sie dabei wieder nur Fachwissen und -können, das Studierende bereits beim Eintritt ins Studium mitbringen müssen, welches also nicht in anderen Lehrveranstaltungen an der Universität/ETH erworben werden kann."

Eine verdichtete Zusammenstellung der Ergebnisse ist in Tabelle II.7 zu finden. Die Zahlen geben jeweils die Anzahl gleicher Nennungen an[4].

Tab. II.7: Konkret genanntes, vorausgesetztes Fachwissen und -können

Fach	Inhalte
Deutsch	• Beherrschung der Grammatik: 7 Germanistik-Dozierende; 7 Dozierende weiterer, je verschiedener Studienfächer. • Sprachbeherrschung: 55 Nennungen quer durch alle Studienfächer, dabei nahezu 100% der Dozierenden der Rechtswissenschaften; sehr gute Sprachbeherrschung insbesondere auch in stilistischer Hinsicht: 20 Dozierende aus unterschiedlichen Studienfächern; wissenschaftliche Schreibweise: 2 Dozierende der Biologie. • Syntax (6); Orthographie (8); Interpunktion (3); Semantik (2). • Aus den Reihen der Germanistik-Dozierenden: Grundkenntnis der Sprachgeschichte (1), Überblick über die wichtigsten Epochen der Literaturgeschichte (6), Kenntnisse einiger grundlegender Interpretations- und Textanalysemethoden (6), Lesen eines eigentlichen Literaturkanons (3), Kenntnis literaturwissenschaftlicher Grundbegriffe (2). • Kernbegriffe der Erzählanalyse, Dramatheorie, Lyrik und Literaturgeschichte: 2 Dozierende der Anglistik; Grundlagen der Literaturinterpretation: 1 Dozierender des Studienfachs Anglistik. • Von Germanistik-Dozierenden weiterhin genannt: Grundlagen der Linguistik (1); allgemeine Erfahrungen mit schwierigen oder literarischen Texten (3); Grundlagen der Sprachstruktur des Deutschen und grammatikalische Grundbegriffe (2); Wissen um Textsorten, Medien- und Kommunikationsformen (1); Kenntnisse der wichtigsten Regeln abendländischer Poetik (1); lesetheoretische Grundmodelle in Wort und Tat (1); Textwissen (1) und genaues Lesen (1), hier zusätzlich 1 Nennung aus dem Studienfach Rechtswissenschaft.
Mathematik	• Infinitesimalrechnung: 5 Dozierende der Wirtschaftswissenschaften, 3 der Informatik, 7 der Biologie, 3 der Medizin und 6 Dozierende weiterer unterschiedlicher Studiengänge; nur Differenzialrechnung: 4 Dozierende unterschiedlicher Studiengänge; nur Integralrechnung: 1 Dozierender der Kommunikationswissenschaft. • Vertiefte Kenntnisse der Algebra: 3 Dozierende der Wirtschaftswissenschaften, 11 Dozierende anderer Studiengänge; Exponential- und Logarithmusfunktionen: 4 Dozierende der Biologie; Wahrscheinlichkeitsrechnung: 4 Dozierende der Psychologie, 7 weiterer Studiengänge; Statistik und deren Interpretation: 4 Dozierende der Biologie, 7 Dozierende anderer Studienfächer. • Mengenlehre (3); Logik (6); Planimetrie (2); Vektorgeometrie (7); Trigonometrie (6); Wissen über komplexe Zahlen (2); Funktionenlehre (7); Arithmetik (6, davon 3 Dozierende der Informatik); Zahlensysteme (2). • Basiskenntnisse in den Bereichen Spieltheorie (1); Folgen und Reihen (2); Polynome (1); Kombinatorik (2); Binomialverteilung (1); Differenzialgleichungen (1); Beweiskonzepte (1); Wellenfunktionen (1).

[4] Falls eine weitere Angabe zur Zahl fehlt, stammen die Nennungen von Dozierenden verschiedener Studienfächern.

Fach	Inhalte
Biologie	• Grundbegriffe der Evolution (1); anatomischer oder physiologischer Aufbau des Menschen: 2 Dozierende, jeweils aus nichtnaturwissenschaftlichen Studienrichtungen. • Studienrichtung Medizin: Vererbungslehre, Molekulargenetik, Aufbau pflanzlicher wie tierischer Zellen, Stoffkreisläufe, Ökologie, Elementares zur Mikrobiologie und Evolutionslehre: 1 Nennung; Grundzusammenhänge der Molekularbiologie, Molekulargenetik, Biochemie: 1 Nennung. • Nennungen aus dem Studienfach Biologie: Grundlage der Zellteilung (2); Photosynthese (2); Botanik (1); Zoologie (1); Aufbau der Zelle (1); Grundbegriffe der Ökologie (4); Kenntnisse der Evolutionsbiologie (2); Evolutionsgenetik als Prinzip (1); klassische Genetik (4); molekulare Genetik (4); Molekularbiologie (2); Neurobiologie (1); Physiologie (3); Organismenkenntnisse (1); Biodiversitätskrise (1); Artkonzepte (1); Stoffwechsel (1); Atmung (1); Ernährung (1); Fortpflanzung (1); Membranen (1); Zellbiologie (2); Evolution (5); Proteinbiochemie (1); Grundlagen in Anatomie (1).
Griechisch	• Grundkenntnisse: 3 Dozierende aus den Fachbereichen Germanistik und Geschichte. • Beherrschung des griechischen Alphabets: 1Dozierender der Kommunikationswissenschaft.
Latein	• Grundkenntnisse (5); Kenntnisse zur Vorbereitung sprachlicher Strukturanalysen (1); Grundwortschatz (2). • Vertiefte Kenntnisse zur Bearbeitung von Originalquellen: 2 Dozierende der Geschichte.
Französisch	• Sprachbeherrschung: 4 Dozierende der Geschichte, 1 Dozierender der Rechtswissenschaft; Kenntnisse zur Verarbeitung von Fachtexten: 7 Dozierende der Rechtswissenschaften, 5 der Geschichte sowie 1 Dozierender der Politikwissenschaft; Textverständnis (3).
Englisch	• Hohes Niveau der Sprachbeherrschung (Sprachenportfolio mindestens C1): alle befragten Dozierenden der Anglistik • Fähigkeit des Hörverstehens, Kenntnisse der verschiedenen Sprachebenen, Kernbegriffe aus der Erzählanalyse, Dramatheorie, Literaturgeschichte: jeweils 1 (unterschiedlicher) Dozierender der Anglistik. • Sprachkenntnisse auf zur Verarbeitung von Fachliteratur erforderlichem Niveau in Wort wie in Schrift (54).
Italienisch/ Spanisch	• Italienischkenntnisse zur Verarbeitung von Urteilen: 3 Dozierende der Rechtswissenschaften. • passive Italienischkenntnisse: 1 Dozierender aus der Geschichte.
Geschichte	• Breites Basiswissen: 5 Dozierende der Germanistik, 4 der Anglistik, Geschichte und der Rechtswissenschaften. • Allgemeinen Überblick über die Schweizer und die Europäische Geschichte der Neuzeit: 5 Dozierende der Geschichte, 10 Dozierende aus verschiedenen Studienfächern, vertiefte Kenntnisse in der Epoche des MA: 1 Dozierender der Germanistik, Ausdehnung auf den nordamerikanischen Kontinent: 1 Dozierender der Anglistik. • Kenntnisse der jüngeren politisch-sozialen Geschichte (5). • Grundbegriffe, ein gewisses Jahreszahlengerüst und der Umgang mit Quellen: 4 Dozierende der Geschichte; Wissen aus dem Epochenspannungsbereich von Antike bis zur Gegenwart: je 1 Dozierender der Geschichte und der Politikwissenschaft.
Wirtschaft und Recht	• Grundlegende Kenntnisse (Wirtschaftskreislauf, Angebot/Nachfrage, Recht, Moral): 3 Dozierende der Geschichte, 3 der Wirtschaftswissenschaften, 4 Dozierende weiterer Studiengänge; vertieftes Verständnis über den Aufbau des öffentlichen und privaten Rechts: 3 Dozierende des Rechts. • Kenntnisse über das Wirtschaftssystem: 1 Dozierender der Soziologie; Wissen über Wertschriften, Zins und Zinseszinsen sowie die begriffliche Gegenüberstellung von Kosten und Nutzen: 1 Dozierender der Informatik.
Geographie	• Kenntnisse zur Orientierung im Raum (z. B. Mittelmeerraum): 3 Dozierende der Geschichte; Kenntnisse über humangeographische Aspekte (Veränderung von Lebensräumen in historischer Dimension, z. B. Raumplanung): 1 Dozierender der Geschichte. • Kenntnisse des Klimas, über die Klimazonen der Kontinente, der Meteorologie und der Plattentektonik:2 Dozierende der Biologie.
Chemie	• Vertiefte grundlegende Kenntnisse der Chemie (allg. Chemie, Atome, chemische Bindungen, VSEPR-Modell, Gleichgewichte, PSE, Chemie der Elemente, organische Chemie, chemische Bindungen, Atomaufbau, Moleküle, Bindungskräfte, Salze, Metalle, chemische Reaktionen, stöchiometrische Gleichungen, Thermodynamik, Säure und Basen, Redoxreaktionen, Formelgewicht, Periodensystem, Atomgewicht, Ion, Mol, Oxidationszahl, funktionelle Gruppen wie Alkohole, Carbonsäuren, Ketone, Aldehyde, Amine, Amide): 2 Dozierende der Medizin und 6 Dozierende der Biologie.

Fach	Inhalte
Physik	• Kenntnisse der Akustik und Wellenlehre: je 1 Dozierender der Germanistik und der Anglistik. • Kenntnisse zum Einschätzen von Grössenordnungen, zu SI-Einheiten, Grundlagen der Mechanik, Optik, Akustik, Thermodynamik-Wärmelehre und der Elektrizitätslehre: 7 Dozierende der Biologie, 3 Dozierende der Physik und 4 weitere Dozierende naturwissenschaftlich geprägter Studienfächer.
Philosophie	• Grundlegende Kenntnisse: 3 Dozierende der Germanistik, 7 Dozierende weiterer unterschiedlicher Studienfächer; Wissen über die wichtigsten Erzählungen der Bibel: 3 Dozierende der Geschichte, 1 Dozierender der Pädagogik. • Spezifische Denkschulen (insbesondere Sokrates, Platon, Descartes, Hume, Leibniz, Kant, Humboldt, Hegel, Marx, Nietzsche, Wittgenstein, Sartre): 4 Nennungen. • Kenntnisse der Philosophie und Philosophiegeschichte des 19. und 20. Jhdt.: 1 Dozierender der Architektur; grundlegende Kenntnisse der Argumentationstechniken: 1 Dozierender der Politikwissenschaft. • Überblick über die Ideengeschichte: je 1 Dozierender der Rechtswissenschaften und der Architektur. • Ausgewählte klassische Pädagogen, Befunde aus Familien und Schulforschung, entwicklungspsychologische Grundlagen und das Wissen um die Abgrenzung der Pädagogik zu anderen Fächern: 2 Dozierende der Pädagogik.
Informatik	• Benutzerwissen zur Anwendung gebräuchlicher Anwenderpakete (MS-Office-Programme): 45 Nennungen aus fast allen Studiengängen. • Produktives Arbeiten mit Datenbanken (und nicht nur mit Google) und Algorithmen (7); eigentliches Programmieren (4); Umgang mit einem Betriebssystem: 1 Dozierender der Informatik.
Musik	• Es wurde kein Eingangswissen und -können erwartet.
Bildnerisches Gestalten	• Grundlegende Kenntnisse der Kunstgeschichte:1 Dozierender der Anglistik; Kenntnisse über die Analyse visueller Zeichen, Bilder und Photographien: 1 weiterer Dozierender der Anglistik; Fähigkeit, einen Organismus bildnerisch darstellen zu können: 1 Dozierender der Biologie; Fähigkeit, aus der eigenen Vorstellung zeichnen zu können: 1 Dozierender der Architektur.

Bei den Antworten auf diese Frage zeigt sich insbesondere, dass Abweichungen vom curricularen Verständnis des jeweiligen Faches am Gymnasium bestehen können. Wenn Dozierende von der grossen Bedeutung von Deutsch sprechen, meinen sie vorwiegend „Sprachbeherrschung“, Grammatik, Syntax u. ä. Eigentliche Kerninhalte der Germanistik wie Sprach- oder Literaturgeschichte, sprachwissenschaftliche Interpretations- und Textanalysemethoden, Grundlagen der Linguistik, Dramatheorie usw. werden nur von einigen der Germanistinnen und Germanisten sowie einem Anglisten erwartet. Es verbergen sich also hinter dem Begriff „Fach Deutsch“ unterschiedliche Vorstellungen hinsichtlich der Fachinhalte. Wenn Deutsch allgemein als wichtigstes Maturafach eingeschätzt wird, ist somit hauptsächlich deutsche Sprachkompetenz gemeint. Ähnlich verhält es sich beim Fach Englisch, bei dem die Fähigkeit zur Bearbeitung englischsprachiger Fachliteratur (und nicht etwa literaturgeschichtliche Kompetenz) im Mittelpunkt steht.

3.3.3 Aktuell festgestellte Wissens- und Könnenslücken (Frage 3)

Mittels der folgenden weiteren Frage wurden die Dozierenden nach aktuell festgestellten Lücken im zuvor als wichtig bezeichneten Wissen und Können bei Studienanfängerinnen und -anfängern befragt.

„Über welches in der Frage 2 bezeichnete fachliche Wissen und Können verfügen die neu eintretenden Studierenden meistens nicht (mehr), um Ihrer Lehrveranstaltung folgen zu können?“

Eine wieder verdichtete Zusammenstellung der Ergebnisse mit der jeweiligen Anzahl[5] Nennungen ist in Tabelle II.8 zu finden.

Tab. II.8: Konkret genannte Lücken beim vorausgesetzten Fachwissen und Können

Fach	Inhalte
Deutsch	• Sprachbeherrschung (teilweise ohne genauere Erklärung, welche Teilaspekte der Sprache nicht beherrscht werden): 11 Dozierende der Rechtswissenschaft, 3 Dozierende der Geschichte, 6 Dozierende, die den nicht sprachlich orientierten Studienrichtungen angehören. • Grammatik: 5 Dozierende der Germanistik, 6 Dozierender der Geschichte, 2 weitere Dozierende unterschiedlicher Studiengänge; Orthographie und Rechtschreibung: 4 Dozierende der Germanistik, 4 Dozierende unterschiedlicher Studienfächer; Syntax: 3 Dozierende der Geschichte, 1 Dozierender der Germanistik; Interpunktion: 4 Dozierende der Geschichte, 1 Dozierender der Germanistik. • Kenntnisse der Literaturgeschichte: 5 Dozierende der Germanistik; Wissen zur Interpretation von literarischen Texten: 3 Dozierende unterschiedlicher Studienfächer; Kenntnisse über Linguistik: je 1 Dozierender der Germanistik und der Anglistik. • Verfassen wissenschaftlicher Texte, Umgang mit wissenschaftlichen Texten: je 1 Dozierender der Politikwissenschaft und der Anglistik. • Literaturanalytisches Begriffsinventar: 3 Dozierende der Germanistik; Wissen um Etymologie:1 Dozierender der Germanistik. • Erarbeitung komplexer Texte: 1 Dozierender der Wirtschaftswissenschaften. • „Struktureller Analphabetismus“: 1 Dozierender der Rechtswissenschaften.
Mathematik	• Statistik, Wahrscheinlichkeitsrechnung (7). • Infinitesimalrechnung (5); Integralrechnung (4). • Komplexe Zahlen: je 1 Dozierender der Informatik und der Physik; Vektorgeometrie: 4 Dozierende unterschiedlicher Studiengänge. • Exponential- und Logarithmusfunktionen (3). • Algebra: je 1 Dozierender der Wirtschaftwissenschaften und der Informatik. • Aussagenlogik (3); Beweise: 1 Dozierender der Informatik. • Arithmetik: je 1 Dozierender der Informatik und der Medizin. • Trigonometrie: 1 Dozierender der Architektur. • Funktionenlehre: je 1 Dozierender der Wirtschaftswissenschaft und der Physik. • Binäres Zahlensystem: 1 Dozierender der Physik. • Umgang mit Folgen und Reihen: je 1 Dozierender der Physik und der Architektur. • Winkelfunktionen: 1 Dozierender der Medizin. • Quadratische Gleichungen und Strahlensätze: 1 Dozierender der Architektur. • Zahlentheorie und analytische Geometrie: 1 Dozierender der Informatik.
Biologie	• Anwendung anatomischer Kenntnisse auf die Sprachwissenschaft: 1 Dozierender der Anglistik. • Basiswissen hinsichtlich vegetativem Nervensystem, Kognition, Neuroanatomie: 1 Dozierender der Pädagogik.

[5] Falls eine weitere Angabe zur Zahl fehlt, stammen die Nennungen von Dozierenden verschiedener Studienfächern.

Fach	Inhalte
Griechisch	• Grundkenntnisse und Wissen: je 1 Dozierender der Germanistik und der Geschichte.
Latein	• Grundkenntnisse (5).
Französisch	• Kenntnisse zur Bearbeitung von Fachliteratur: 4 Dozierende der Geschichte, 3 Dozierende der Rechtswissenschaft, 1 Dozierender der Politikwissenschaft; Kenntnisse eines fachlichen Grundwortschatzes: 1 Dozierender der Rechtswissenschaft.
Englisch	• Niveau der Sprachbeherrschung und Verständnis sprachlicher Nuancen: 6 Dozierende der Anglistik; Grammatikkenntnisse und Wortschatz: 1 Dozierender der Anglistik; Sprachbeherrschung im Umgang mit wissenschaftlichen Texten: 2 Dozierende der Anglistik; linguistische Kenntnisse: 1 Dozierender der Anglistik. • „advanced level" bei Eintritt: 1 Dozierender der Anglistik. • Übung im Umgang mit komplexeren Texten der Fachliteratur, Wortschatz und differenzierteres Sprachverständnis: 5 Dozierende der Pädagogik, 4 Dozierende der Geschichte, 8 weitere Dozierende unterschiedlicher Studienrichtungen.
Italienisch/ Spanisch	• Grundkenntnisse (3).
Geschichte	• differenziertes Überblickswissen über die letzten 250 Jahre sowie die Zeit davor (Chronologie, Zusammenhänge): 8 Dozierende der Geschichte, 5 Dozierende weiterer unterschiedlicher Studienrichtungen; breites Wissen der Schweizer Geschichte: 1 Dozierender der Geschichte. • Bewusstsein für Historizität: je 1 Dozierender der Germanistik und der Geschichte. • Nicht klischeehafte Aufnahme der mittelalterlichen Geschichte: 2 Dozierende der Germanistik. • Kenntnisse über die Antike (3). • Kenntnisse zur amerikanischen Geschichte und der Kolonialzeit: 1 Dozierender der Anglistik. • Kenntnisse in Staatskunde (3).
Wirtschaft und Recht	• Allgemeinverständnis der Wirtschaft und ihrer grundlegenden Struktur: 3 Dozierende der Geschichte. • Grundlagen des Rechts: 1 Dozierender der Pädagogik. • Wissen über Kosten, Zinsen, Wertpapiere: 1 Dozierender der Informatik.
Geographie	• Allgemeinwissen: 1 Dozierender der Geschichte.
Chemie	• Aktuelles Wissen, naturwissenschaftliche Vorbildung: 1 Dozierender der Medizin.
Physik	• Akustik: je 1 Dozierender der Germanistik und der Anglistik. • Umgang mit Masseinheiten, Kenntnisse über Druck, Gewichtskraft und momentane Geschwindigkeiten: je 1 Dozierender der Informatik und der Architektur.
Philosophie	• Grundlegende Kenntnisse, Wissen über biblische Geschichten: 5 Dozierende der Geschichte, 3 weitere Dozierende unterschiedlicher Studienrichtungen. • Philosophie der Antike: 1 Dozierender der Pädagogik. • Grundkenntnisse über Argumentationstechniken: 1 Dozierender der Politikwissenschaften.
Informatik	• Benutzerwissen (6). • Umgang mit Algorithmen: 1 Dozierender der Anglistik. • Mut zum Programmieren bzw. Kenntnisse über Programmiersprachen: je 1 Dozierender der Soziologie und der Informatik.
Musik	• Es werden keine Mängel festgestellt.
Bildnerisches Gestalten	• Kenntnisse der Bildanalyse: 1 Dozierender der Anglistik.

Es zeigt sich, dass in fast allen Fächern Lücken im Eingangswissen und -können der Studierenden genannt wurden; herausragend viele in Erstsprache und Mathematik. In Erstsprache sind es dabei wieder im Wesentlichen Komponenten der „Sprachbeherrschung“, deren „Dürftigkeit“ vor allem nicht Germanistik Dozierende monierten. Lücken bei den eigentlichen Kerninhalten der Germanistik wurden wieder nur von einigen der Germanistinnen und Germanisten sowie einem Anglisten festgestellt.

3.3.4 Wichtigkeit überfachlicher Kompetenzen für die Lehrveranstaltung (Frage 4)

Die Dozierenden sollten in einer weiteren Frage auf einer siebenstufigen Skala einschätzen, für wie wichtig sie vorgelegte überfachliche Kompetenzen für das Verständnis und den erfolgreichen Besuch ihrer Veranstaltungen erachten. Die Frage lautete wie folgt:

„Welches sind weitere wichtige, überfachliche Kompetenzen, über welche Studierende bereits beim Studienbeginn verfügen müssen, damit sie in der Lage sind, Ihrer Lehrveranstaltung zu folgen?“

Die aus der Tabelle II.9 in der ersten Kolonne ersichtlichen überfachlichen Kompetenzen wurden im Wesentlichen aus der Befragung von Notter & Arnold (2003 und 2006) übernommen (siehe Kapitel I, Abschnitt 3), teilweise aber auch aufgrund weiterer, in der Literatur genannter Komponenten einer allgemeinen Studierfähigkeit und interner Diskussionen innerhalb des EVAMAR-Teams (siehe Kapitel I, Abschnitt 5.4) modifiziert und ergänzt.

Huber (1998, S. 158) z. B. beschreibt folgende allgemeine Kompetenzen und Haltungen:

- (gegenstandsbezogen) Studentenmotivation, Lernbereitschaft, Fachinteresse, Neugier;
- (zukunftsbezogen) Orientierung, Gewissheit in der Berufsperspektive und in der Fachwahl (oder bewusstes vorläufiges Absehen von ihr);
- (sozialitätsbezogen) Kontakt-, Kommunikations- und Kooperationsbereitschaft und -fähigkeit;
- (selbstbezogen) realistisches Selbstbewusstsein, mehr noch: Selbstvertrauen und genügende Ich-Stärke, sich auf Infrage-Stellungen einzulassen, entwickelte oder sich entwickelnde Identität, und daraus resultierend:
- Selbstständigkeit: inhaltlich (Themenfindung), organisatorisch (Raum, Zeit, Koordination), methodisch (wissenschaftliche Arbeitstechniken), lebenspraktisch (Wohnung, Jobs).“

Schüpbach et al. (2006, 156) haben als Komponenten einer Handlungskompetenz im Studium des Hochschulfachs Englisch die folgenden Elemente ermittelt: selbstständiger Erwerb englischer Sprachpraxis, Informationssuche, Teamarbeit, selbstorganisiertes Lernen, Ressourcennutzung, Prüfungsvorbereitung, Freizeiteinschränkung, selbstständiges Arbeiten, aktive Teilnahme an Veranstaltungen. Ähnliche Ergebnisse resultieren bei Wetzenstein (2004) aus einer Anforderungsanalyse für das Psychologiestudium, hier rangieren die Personeneigenschaften und Fähigkeiten zur Selbstorganisation, die allgemeine Sprachkompetenz, Leistungsmotivation, logisches Denken, kritisches Denken, Intelligenz, Belastbarkeit, Gedächtnisleistung sowie Interesse an Unbekanntem/Offenheit und Abstraktes Denken auf den vorderen Plätzen.

Es handelt sich bei den im Fragebogen verwendeten Teilkompetenzen vorwiegend um theoretisch und empirisch noch wenig gut fundierte Konstrukte. Der Einbezug in die Dozierendenbefragung ermöglicht aber den Vergleich mit der zuvor erwähnten Studie, in der die empirischen Daten zur Bedeutung der Teilkompetenzen nur als Einschätzung der Studierenden, nicht aber aus Sicht der Lehrenden an Universitäten erhoben wurden.

Die Ergebnisse finden sich in Tabelle II.9, sortiert nach dem Mittelwert (je höher, desto wichtiger).

Tab. II.9: Wichtigkeit überfachlicher Kompetenzen für die eigene Lehrveranstaltung

Kompetenzen	Mittelwert	Standard-abweichung	Wichtigkeit aus Sicht Studieren-der*
Selbstständiges Lernen	6.48	.828	1
Verantwortung für eigenes Lernen und Arbeiten übernehmen	6.40	.872	3
Selbstständiges Arbeiten	6.38	.882	2
Kritisches Denken	6.33	1.003	7
Problemlösefähigkeit	6.11	1.243	6
Schriftliche Ausdrucksfähigkeit	5.89	1.407	8
Zeit effizient einteilen und Prioritäten setzen	5.85	1.104	4
Mit Belastungen umgehen	5.59	1.339	5
Hörverstehen im Zusammenhang mit Vorlesungen	5.55	1.420	
Mündliche Ausdrucksfähigkeit	5.42	1.715	10
Umfangreiche Prüfungen vorbereiten	5.25	1.586	
Entwickeln neuer Ideen	5.20	1.374	11
In kurzer Zeit viel Lernstoff verarbeiten	5.14	1.405	
Recherchierfähigkeit	5.02	1.648	
Im Team arbeiten	4.74	1.544	9
Vor Publikum auftreten	4.16	1.819	12

N max. = 140

* Wichtigkeitsrang in den ersten Semestern aus Sicht Studierender gemäss Notter & Arnold (ohne Kategorie „Textverständnis") (2006, S. 52).

Alle vorgelegten Kompetenzen wurden also als sehr wichtig (nahe bei der Stufe 7) bis mindestens „mittelwichtig" (leicht über Stufe 4) eingestuft. Die Reihenfolge stimmt in der Tendenz erstaunlich gut überein mit der Wertung der Wichtigkeit dieser überfachlichen Kompetenzen durch die Studierenden in den Studien von Notter & Arnold (2003 und 2006). Ein Vergleich der Bewertung nach Studienfächern bestätigt diese Bedeutungszumessungen im Grossen und Ganzen. Einzig „Vor Publikum auftreten" fällt bei vielen Studienfächern unter die Stufe 4.

Die Dozierenden konnten in einem offenen Antwortteil auch weitere, in ihrer Einschätzung wichtige überfachliche Kompetenzen angeben. Während einige der genannten Fähigkeiten den vorgegebenen ähneln oder sich direkt aus ihnen ableiten, können andere als sinnvolle Ergänzung betrachtet werden: Konzentrationsfähigkeit, Ausdauer, Präzision, logisches, vernetztes, abstraktes (z. B. dreidimensionales Vorstellungsvermögen) und analytisches Denken, assoziatives Denken, Fähigkeit zum Transfer, realistische Selbsteinschätzung, Selbstkritik, Argumentieren, Diskussionsfähigkeit, Lesekompetenz.

3.3.5 Festgestellte überfachliche Wissens- und Könnensdefizite (Frage 5)

Auch bei den überfachlichen Kompetenzen wurde nach festgestellten Defiziten gefragt, und zwar mittels folgender Formulierung:

„Über welches in der Frage 4 als sehr wichtig bis wichtig (Kästchen 1-3) bezeichnete überfachliche Wissen und Können verfügen die neu eintretenden Studierenden meistens nicht (mehr), um Ihrer Lehrveranstaltung folgen zu können?“

Die Antworten auf diese offene Frage wurden den in Frage 4 vorgegebenen Kategorien zugeordnet und quantifiziert (siehe Tabelle II.10). Es fällt der hohe Anteil der Nennung von Defiziten bei der schriftlichen Ausdrucksfähigkeit, beim kritischen Denken, selbstständigen Arbeiten und selbstständigen Lernen auf. Gerade diese überfachlichen Kompetenzen belegen auf der Wichtigkeitsskala die vorderen Plätze. Dieser Zusammenhang zwischen Bedeutsamkeit der Kompetenz und der Angabe eines Defizits (je bedeutsamer, desto eher defizitär) lässt sich über die gesamte Tabelle feststellen. Er stimmt zudem recht gut überein mit den entsprechenden Selbsteinschätzungen der Studierenden in der Studie von Notter & Arnold (2006).

Tab. II.10: Defizite in als wichtig erachtetem überfachlichem Wissen und Können

Kompetenzen	kein Manko	Kompetenz wichtig, aber lückenhaft	Mittelwert Wichtigkeit
Schriftliche Ausdrucksfähigkeit	57.6%	42.4%	5.89
Kritisches Denken	65.0%	35.0%	6.33
Selbstständiges Arbeiten	70.0%	30.0%	6.38
Selbstständiges Lernen	74.0%	26.0%	6.48
Mündliche Ausdrucksfähigkeit	79.0%	21.0%	5.42
Verantwortung für eigenes Lernen und Arbeiten übernehmen	79.4%	20.6%	6.40
Problemlösefähigkeit	80.0%	20.0%	6.11
Recherchierfähigkeit	81.0%	19.0%	5.02
Zeit effizient einteilen und Prioritäten setzen	84.7%	15.3%	5.85
Mit Belastungen umgehen	84.8%	15.2%	5.59
Hörverstehen im Zusammenhang mit Vorlesungen	88.9%	11.1%	5.55
Entwickeln neuer Ideen	91.0%	9.0%	5.20
Umfangreiche Prüfungen vorbereiten	91.8%	8.2%	5.25
In kurzer Zeit viel Lernstoff verarbeiten	92.0%	8.0%	5.14
Vor Publikum auftreten	97.0%	3.0%	4.16
Im Team arbeiten	97.0%	3.0%	4.74

N max. = 100

3.3.6 Wünschbares zusätzliches Eingangswissen und -können

Mit der folgenden letzten Frage des Fragebogens wurden die Dozierenden nach weiterem Wissen und Können gefragt, welches sie sich zusätzlich – über das für ihre eigene Lehrveranstaltung unabdingbare hinaus – als Inhalt gymnasialer Curricula wünschen würden.

„Wenn Sie selbst die Bildungsziele des Gymnasiums festlegen könnten, welches im Vergleich zu Ihren Antworten zu den Fragen 2 und 4 zusätzliche fachliche und überfachliche Wissen und Können sollten Studierende bereits im Gymnasium im Hinblick auf Ihre Lehrveranstaltung erworben haben? (Bitte auch begründen)“

Die Ergebnisse sind in Tabelle II.11 zusammengefasst[6]. Sie sind insofern nur teilweise konsistent, als sie Inhalte einschliessen, die bereits jetzt an den meisten Gymnasien unterrichtet werden (besonders gut ersichtlich bei Mathematik). Das lässt vermuten, dass nicht wenige Dozierende unzureichend darüber informiert sind, was an den Gymnasien effektiv unterrichtet wird.

Tab. II.11: Weiteres, wünschbares Wissen und Können bei Studienbeginn

Fach	Inhalte
Deutsch	• Begriffe zur Text- und Literaturanalyse und Textinterpretation: 4 Dozierende der Germanistik, 1 Dozierender der Anglistik. • Textinterpretation im Spannungsfeld zu eigener Biographie und Werthaltung, Textbeobachtung: 1 Dozierender der Germanistik. • Kenntnisse der Literaturgeschichte: je 1 Dozierender der Germanistik und der Pädagogik. • Wissen über die deutsche Literatur und deren Parallelen zur englischen Literaturgeschichte: 1 Dozierender der Anglistik. • Rhetorik: 1 Dozierender der Germanistik. • historische Kontextualisierung von Texten, Bewusstsein historischer Sprachstufen und Sprachentwicklung: 1 Dozierender der Germanistik. • Deutsch als Instrument beruflicher Kommunikation: 1 Dozierender der Rechtswissenschaften. • Kenntnisse der Linguistik, Jugend- und Mediensprache, Sprachwandel, Zeichensysteme, Sprachstrukturen: je 1 Dozierender der Germanistik und der Anglistik. • Sprachbeherrschung (3); Ausdrucksfähigkeit: je 1 Dozierender der Politikwissenschaften und der Rechtswissenschaften. • Verfassen verschiedener Textsorten: 1 Dozierender der Germanistik; strukturiertes, wissenschaftlich, fachorientiertes Schreiben: 1 Dozierender der Informatik. • Grammatik, Syntax, Textwissen, Orthographie, Interpunktion, angstfreies Herangehen an alte Texte: 1 Dozierender der Germanistik.
Mathematik	• Wahrscheinlichkeitsrechnung: je 1 Dozierender der Psychologie und der Physik; Stochastik: 1 Dozierender der Pädagogik. • Umsetzung der Theoreme in konkrete praktische Fragestellungen: 1 Dozierender der Soziologie. • Einführung des Begriffs der Differenzialgleichung: 1 Dozierender der Informatik. • Induktionsbeweise: 1 Dozierender der Informatik; Prinzipien der Beweisführung: 1 Dozierender der Physik. • intuitives Verständnis geometrischer und analytischer Konzepte: 1 Dozierender der Physik. • algorithmisches Denken: 1 Dozierender der Physik. • Verständnis von Statistik, Daten prägnant darstellen, interpretieren und analysieren: 1 Dozierender der Biologie. • komplexe Zahlen: 1 Dozierender der Biologie. • einfache Differenzialgleichungen: 1 Dozierender der Biologie. • Vektorprodukt, Kreisgleichung, Logarithmusfunktionen: 1 Dozierender der Architektur. • Verstehen funktionaler Zusammenhänge aus graphischen Darstellungen: 1 Dozierender der Architektur.
Biologie	• biologische Grundlagen von Verhalten, Neuro- und Sinnesphysiologie: 1 Dozierender der Psychologie. • mehr ökologische Fragestellungen behandeln, Wissen um aktuelle ökologische Probleme der Schweiz und der Welt (3). • historische Aspekte der Biologie, vergleichende Biologie: 1 Dozierender der Biologie. • Organismenkenntnisse im Zusammenhang mit der Lehre der diversitätsbiologischen Zusammenhänge:je 1 Dozierender der Biologie und der Medizin. • Stoffkreisläufe, Evolution: 1 Dozierender der Medizin.
Griechisch	• Kenntnisse des griechischen Alphabets: 1 Dozierender der Geschichte. • fundierte Kenntnisse in alten Sprachen: 1 Dozierender der Geschichte.

[6] Falls eine weitere Angabe zur Zahl fehlt, stammen die Nennungen von Dozierenden verschiedener Studienfächern.

Fach	Inhalte
Latein	• minimale Kenntnisse des Grundwortschatzes: 1 Dozierender der Germanistik. • Lateinkenntnisse sollten obligatorisch sein: 3 Dozierende der Geschichte. • zur Schulung des Sprachempfindens im Deutschen, bessere Beherrschung der medizinischen Begriffe: 1 Dozierender der Medizin.
Französisch	• müheloses Verstehen von französischen Texten und Debatten: 1 Dozierender der Politikwissenschaften.
Englisch	• Englisch auf advanced level – mindestens C2: 2 Dozierende der Anglistik. • Bearbeitung auch schwieriger Textsorten (Wissenschaftssprache): 1 Dozierender der Anglistik. • Grundwissen über Geschichte zur Einordnung in die englische Literaturgeschichte: 1 Dozierender der Anglistik. • höhere Sprachkompetenz: 1 Dozierender der Anglistik. • gute Kenntnisse (3).
Italienisch/ Spanisch	• Lesen von Fachliteratur in Italienisch: 1 Dozierender der Geschichte.
Geschichte	• Einsichten in die Prozesse der Geschichte: 1 Dozierender der Germanistik; zusammen mit Philosophie unterrichten: 1 Dozierender der Germanistik; geschichtliches Wissen: 1 Dozierender der Anglistik. • rudimentäre Kenntnisse der Ereignisstrukturen in der Schweiz und in Europa: 1 Dozierender der Geschichte; harte Fakten-Kenntnisse und das basale Handwerkszeug eines Historikers: 1 Dozierender der Geschichte; Kenntnis der wichtigsten Epochen der Weltgeschichte mit Daten: 1 Dozierender der Geschichte; Allgemeinwissen über soziale, kulturelle und politische Grundlagen der eigenen ‚Lebenswelt': 1 Dozierender der Geschichte; breiter Überblick über Ereignisse und Fragestellungen der europäischen Geschichte: 2 Dozierende der Geschichte; Geschichte findet auch noch vor 1789 statt: 2 Dozierende der Geschichte; Kenntnis der weltgeschichtlichen Zusammenhänge nach der franz. Revolution: 1 Dozierender der Politikwissenschaften; geschichtlicher Überblick über die letzten 300 Jahre: 1 Dozierender der Rechtswissenschaften. • dreiteiliger Politikbegriff: Form (polity), Prozeduren (politics), Inhalte (policy): 1 Dozierender der Politikwissenschaften. • Orientierungssinn der Auseinandersetzung mit Geschichte als Basis zur Beurteilung der Gegenwart: 1 Dozierender der Politikwissenschaften. • elementare Grundlagen der Staatskunde (Aufbau des Bundes, Bundesverfassung, Staatsgewalten): 1 Dozierender der Rechtswissenschaften. • Kultur- und Technikgeschichte: 2 Dozierende der Architektur. • Geschichte der Medien: 1 Dozierender der Kommunikationswissenschaften.
Wirtschaft und Recht	• ökonomische Kenntnisse über Medien: 2 Dozierende der Kommunikationswissenschaften; zusätzlich rechtliche Kenntnisse über Medien: 1 Dozierender der Kommunikationswissenschaften. • Mikroökonomie in den Grundzügen und in Modellen: 1 Dozierender der Wirtschaftswissenschaften. • Grundlagen des HRM: 1 Dozierender der Wirtschaftswissenschaften. • Recht: wichtigste rechtliche Aspekte für das Alltagsleben: 1 Dozierender der Wirtschaftswissenschaften.
Geographie	• Ethnologie: 1 Dozierender der Geschichte.
Chemie	• Chemie als obligatorisches Schulfach setzen: 1 Dozierender der Medizin; Abbau von Vorurteilen gegenüber dem Fach (Umwelt): 1 Dozierender der Medizin.
Physik	• keine Einträge.
Pädagogik, Psychologie, Philosophie	• Philosophie zusammen mit Geschichte lehren: 1 Dozierender der Germanistik. • Grundkenntnisse zur Einordnung literaturwissenschaftlicher Phänomene im Englischen: 1 Dozierender der Anglistik. • Vergleichende Religionswissenschaften oder Einsichten in theologisches Denken und Arbeiten: 1 Dozierender der Geschichte; Wissen um Geschichte und Kultur aufgrund der Verbindung von Gesellschaft und Religion: 1 Dozierender der Geschichte. • Philosophie: Erkenntnistheorie als Einführung in die Arbeitsweise der Sozialwissenschaften: 1 Dozierender der Politikwissenschaften. • Psychologie ausdehnen auf Gesellschaftswissenschaften, Verständnis darüber, dass die Wirklichkeit eine von Menschen gemachte ist: 1 Dozierender der Architektur.

Fach	Inhalte
Informatik	• vertiefte Informatik Benutzerkenntnisse: 1 Dozierender der Germanistik; Benutzerwissen Informatik: 1 Dozierender der Pädagogik; gewohnter Umgang mit PC und Systemen: 1 Dozierender der Soziologie. • Einsatz neuer Medien: 1 Dozierender der Germanistik. • grundlegende Kenntnisse über Programmierung: 1 Dozierender der Biologie.
Bildnerisches Gestalten	• kulturelles Wissen: 1 Dozierender der Anglistik.
Musik	• breite Allgemeinbildung: 1 Dozierender der Medizin.
Überfachliche Fähigkeiten	• Lerntechniken und Lernstrategien (3). • Vernetzung gelernter Inhalte (3). • Vorbereitung auf lebenslanges Lernen: 1 Dozierender der Germanistik. • Wissenschaftliches Zitieren (3). • Informationsbeschaffung und Recherchieren (5); kritischer Umgang mit Informationen (4). • Verfassen analytischer Arbeiten, umfassendes Entwickeln und Darstellen eigener Thesen: je 2 Dozierende der Anglistik und der Germanistik. • Pünktlichkeit: 1 Dozierender der Germanistik. • Genauigkeit im Arbeiten (3). • höhere Selbstmotivation: (4); Frustrationstoleranz: je 1 Dozierender der Informatik und der Biologie. • Fähigkeit zu kritischem Denken (6); interdisziplinäres Denken: 2 Dozierende der Anglistik; Unterscheidung von Fakten und Meinungen: 2 Dozierende der Pädagogik; strukturiertes eigenständiges und logisches Denken: 1 Dozierender der Wirtschaftswissenschaften. • mehr Schreibpraxis (3). • vertieftes Lesen und die Fähigkeit, abstrakte komplexe und tiefere Textstrukturen zu erfassen: 4 Dozierende der Geschichte und 4 weitere Dozierende unterschiedlicher Studiengänge. • Interpretation von Bildern: 1 Dozierender der Geschichte. • verbesserte Präsentationstechnik: 3 Dozierende der Geschichte. • selbständiges Arbeiten (3). • selbständiges Lernen (3). • grundsätzliche Auseinandersetzung mit wissenschaftlichem Wissen: 1 Dozierender der Pädagogik; Interesse an wissenschaftlicher Arbeit: 1 Dozierender der Wirtschaftswissenschaften; generelle Freude am Entdecken: 1 Dozierender der Psychologie. • soziale Kompetenzen, Teamarbeit: je 1 Dozierender der Psychologie und der Geschichte. • kommunikative Fähigkeiten: 1 Dozierender der Biologie. • organisatorische Kompetenz: je 1 Dozierender der Architektur und der Biologie. • mündliche und schriftliche Ausdrucksfähigkeit: 1 Dozierender der Architektur.

3.3.7 Schlussfolgerungen

Die Befragungsergebnisse haben im Hinblick auf die Tests fast ausnahmslos keine Korrektur der Erkenntnisse aus den Analysen A1 und A3 erforderlich gemacht, sind aber als eigenständige Ergebnisse des Hauptberichts interessant. Als Ausnahmen dazu sind festzuhalten: In Erstsprache wurde auch ein Bereich „Grammatikalische Kompetenzen/Orthografie“ in den Test für Erstsprache aufgenommen. In einem ergänzenden Fragebogen wurden die Maturandinnen und Maturanden um eine Selbsteinschätzung wichtiger überfachlicher Kompetenzen gebeten. Und schliesslich stützen die Befragungsergebnisse die Wahl von Biologie als naturwissenschaftliches Fach (siehe Kapitel IV).

4 Teilprojekt A2: Analyse von Eingangsprüfungen

4.1 Übersicht

Mit diesem Teilprojekt sollten anhand einer Analyse von Eingangsprüfungen an Schweizer Universitäten ergänzende Hinweise zu den Kompetenzen ermittelt werden, die für ein universitäres Studium vorausgesetzt werden.

Zu diesen Prüfungen gehören zunächst die Aufnahmeprüfungen, die an den meisten Universitäten als Qualifikationsmöglichkeit für den Studienzugang angeboten werden, wenn der Nachweis des Bestehens einer Schweizer Matura fehlt. Im Rahmen der Datensammlung für die Teilprojekte A1 und A3 wurden die angefragten Universitäten auch um Angaben zu allfälligen Aufnahmeprüfungen gebeten. Ergänzend wurden diesbezüglich auch die Internetseiten aller Schweizer Universitäten konsultiert.

Eine erste Übersicht hat ergeben, dass diese Prüfungen in der Regel an das Prüfungsprogramm der Schweizerischen Maturitätsprüfungen gekoppelt, also den Maturaprüfungen an den Gymnasien sehr ähnlich sind. Weil Letztere aber den Untersuchungsgegenstand des Teilprojekts D1 bilden und deshalb die Gefahr zirkulärer Schlüsse bestanden hätte, musste diese Kategorie von Zugangsprüfungen von der weiteren Analyse ausgeschlossen werden.

Als weitere Kategorie von Eingangsprüfungen bieten sich eigentliche Eignungstests an, die sich auf explizite, theoretisch begründete und möglichst auch empirisch validierte Studienerfolgsprognosekonzepte stützen. Aufgrund der Erkundungen an den verschiedenen Universitäten hat sich herausgestellt, dass in diese Kategorie nur der Eignungstest für das Medizinstudium (EMS) sowie die Aufnahmeprüfung für ausländische Studienbewerber an der Universität St. Gallen fallen. Beide Instrumente wurden in der Folge näher untersucht.

4.2 Der Medizineignungstest

Der Eignungstest für das Medizinstudium (EMS) wird in der Schweiz für die auf die beschränkten Studienplätze ausgerichtete Zugangsselektion für das Medizinstudium an verschiedenen Universitäten eingesetzt und enthält Aufgaben, die überfachliche Kompetenzen erfassen. Es handelt sich um einen theoretisch gut fundierten und im Rahmen der Möglichkeiten optimal prognosevaliden Eignungstest. Er stützt sich auf den „Test für medizinische Studiengänge“ (TMS), welcher in Deutschland entwickelt wurde.

Der EMS ist ein auf naturwissenschaftliche Bereiche ausgerichteter Fähigkeitstest. Hänsgen & Spicher (2007, 12) beschreiben ihn zusammenfassend wie folgt: „Es werden keine Fakten abgefragt, sondern Problemstellungen mit den notwendigen Informationen vorgegeben, die zu lösen sind. Intendiert ist, dass es sich um eine Arbeitsprobe für die Studierfähigkeit handelt. Die seinerzeit durchgeführten Anforderungsanalysen für ein Studium der Medizin wurden sehr aufwändig und von verschiedenen Seiten durchgeführt ... . Auf der Grundlage dieser Analysen wurden ca. 50 Bereiche und Aufgabentypen geprüft und die unter vielen Aspekten am besten geeignet erscheinenden 13 für die Erprobungen im Übergangsverfahren des TMS ausgewählt und empirisch erprobt. Von diesen haben vor allem aufgrund der Ergebnisse zur Validität dann 9 Aufgabengruppen in den TMS Eingang gefunden. Der EMS sagt diese Studieneignung gemessen an Massstäben vergleichbarer Tests sehr gut voraus.“ Die Korrelationen der Testergebnisse mit Prüfungsnoten im Studium bewegen sich in einem mittleren Bereich (siehe die konkreten Zahlen in Hänsgen & Spicher, 2001). Die Untertests, für welche diese Zahlen vorliegen, sind die folgenden:

- Quantitative und formale Probleme: Fähigkeit, „im Rahmen medizinischer und naturwissenschaftlicher Fragestellungen mit Zahlen, Grössen, Einheiten und Formeln richtig umzugehen“ (Hänsgen & Spicher, 2007, S. 97).
- Konzentriertes und sorgfältiges Arbeiten: „Fähigkeit, rasch, sorgfältig und konzentriert zu arbeiten“ (S. 101).
- Medizinisch-naturwissenschaftliches Grundverständnis: „Verständnis für Fragen der Medizin und der Naturwissenschaften ... . Der Text könnte so in einem Lehrbuch stehen. Wichtig für das Verständnis dieser Textpassage ist, ob daraus bestimmte logische Schlüsse gezogen werden können. Alle Fakten, die für die Beantwortung der Aufgabe notwendig sind, stehen im Text – spezielles medizinisches Vorwissen ist nicht erforderlich“ (S. 101).

- Muster zuordnen: Fähigkeit, „Ausschnitte in einem komplexen Bild wieder zu erkennen“ (S. 103).
- Diagramme und Tabellen: Fähigkeit, „Diagramme und Tabellen richtig zu analysieren und zu interpretieren. In dieser Form werden während des Studiums zahlreiche Zusammenhänge vermittelt“ (S. 104).
- Textverständnis: Fähigkeit, „umfangreiches und komplexes Textmaterial aufzunehmen und zu verarbeiten“ (S. 98).
- Fakten lernen: Prüft, wie gut man sich Fakten von Patienten einprägen und merken kann (S. 103).
- Schlauchfiguren: räumliches Vorstellungsvermögen: „– eine Funktion, die beispielsweise für das Verständnis von Röntgenbildern wichtig ist. Während des Studiums werden zahlreiche eigentlich dreidimensional zu betrachtende Strukturen und Vorgänge in zweidimensionalen Abbildungen vermittelt“ (S. 97).
- Figuren lernen: prüft, wie gut man sich Einzelheiten von Gegenständen einprägen und merken kann (S. 102).

Beispielaufgaben finden sich z. B. in Hänsgen & Spicher (2007), wo auch der weitere heute verwendete Untertest „Planen und Organisieren“ beschrieben ist.

4.3 Die Aufnahmeprüfung an der HSG

Das an der Universität St. Gallen durchgeführte Aufnahmeprüfungsverfahren für ausländische Studienbewerberinnen und -bewerber wurde vom gleichen Institut entwickelt, welches auch den Test für medizinische Studiengänge (TMS) in Deutschland entwickelt hatte. Im Gegensatz zum EMS gibt es aber noch keine empirischen Untersuchungen über die Prognosevalidität. Ebenso fehlt ein theoretisches Gerüst.

4.4 Schlussfolgerungen

Die für Teilprojekt C gemäss Auftrag vorgesehenen Tests haben den Charakter von standardisierten Schulwissenstests[7], die gemäss Einschätzung von Schallberger (1996) bereits eine hohe Vorhersagegültigkeit besitzen. Da der Einsatz von Aufgaben aus einem eigentlichen Studieneignungstest die Möglichkeit bot, in Ergänzung zu den in EVAMAR II vorgesehenen Fachtests weitere studienrelevante, aber eher wissensunabhängige Kompetenzen zu messen, entschied die Steuergruppe des Projekts EVAMAR II, im Rahmen des Teilprojekts C in der Deutschschweiz auch Aufgaben aus früheren Versionen des EMS vorzulegen. Dieser Testteil war im Auftragskonzept noch nicht vorgesehen.

[7] In Erstsprache haben dann allerdings die weiteren Überlegungen zur Entwicklung eines Kompetenzrasters und eines Tests für eher allgemeinere, fachübergreifende Sprachfähigkeiten geführt (siehe Kapitel III, Abschnitt 2).

III Projektteil B: Modellierung hochschulrelevanten Wissens und Könnens in einzelnen Fächern

1 Allgemeines

In der Planung von EVAMAR II war zunächst vorgesehen, die Ergebnisse der Analysen der Teilprojekte A1-A4 direkt zu Kompetenzmodellen zu verdichten. Dieses Vorhaben erwies sich aber schon bald aus verschiedenen Gründen als zu schwierig. Hauptsächlich deshalb, weil die in Projektteil A untersuchten Lehrtexte und Fachaufgaben zwar direkte Schlüsse auf jenes unabdingbare Wissen und Können zulassen, welches Voraussetzung für deren problemlose Bearbeitung ist, diese Informationen aber nicht ausreichend sind als Grundlagenelemente eines kohärenten, den aktuellen lerntheoretischen Erkenntnissen entsprechenden und empirisch abgesicherten Kompetenzmodells, wie es in der aktuellen theoretischen Diskussion um Bildungsstandards vertreten wird (siehe vor allem bei Weinert, 2001). Zudem decken die ermittelten Sinn- oder Wissenseinheiten und kognitiven Fähigkeiten die Anforderungen der Universitäten nicht flächendeckend ab (siehe im Einzelnen in Kapitel II). Es handelt sich nur um eine Stichprobe, die als ausschliessliche Basis eines vollständigen Kompetenzmodells für Studierfähigkeit noch zu eng wäre.

Es erwies sich deshalb als notwendig, bereits bestehende Kompetenzmodelle in den Fächern auf ihre Tauglichkeit hinsichtlich der Erhebungsziele von EVAMAR II zu untersuchen, sie ganz oder auch nur in Teilen zu übernehmen oder sie weiterzuentwickeln. Im Folgenden werden diese analysierten Modelle daher jeweils kurz beschrieben und in ihrer Relevanz für die Fragestellung in EVAMAR II bewertet. Es handelte sich also beim Teilprojekt B nicht um eine deterministische Deduktion der Kompetenzmodellierung aus den Ergebnissen von Projekt A, sondern um deren Kombination mit bereits elaborierten Modellen. Bei diesen datengestützten Adaptionen ging es im Wesentlichen darum, die eher allgemein und formal gehaltenen Beschreibungen der Bereiche bestehender Kompetenzmodelle anhand des spezifischen Wissens und Könnens, das in Projekt A ermittelt wurde, im Hinblick auf die in EVAMAR II anvisierte Schnittstellenproblematik weiter zu konkretisieren. Die Modellierung erfolgte für die drei in Teilprojekt C getesteten und in D1 untersuchten Fächer Erstsprache, Mathematik und Biologie. Im Sinne der Abgrenzung von dem in der laufenden Diskussion bereits belegten Begriff bzw. Konstrukt des „Kompetenzmodells" mit umfassendem Anspruch im Sinne Weinerts (2001) oder Kliemes et al. (2006) werden die Ergebnisse dieser Modellierung als „Kompetenzraster" bezeichnet. Sie dienen der strukturierenden Beschreibung des Wissens und Könnens in jedem Teilfach in Form einer Matrix mit den Dimensionen „Inhalte" und „Fähigkeiten".

Für die spezielle Ausrichtung des Projekts EVAMAR II galt es im Unterschied zu bereits bestehenden oder sich gegenwärtig im Entwicklungsstadium befindlichen Kompetenzmodellierungen in den verschiedenen Fächern insbesondere die folgenden fünf Aspekte der Fokussierung zu beachten:

1. Fokussierung auf eine relativ hohe Jahrgangsstufe und einen relativ späten Zeitpunkt der Kompetenzerwartung (18- bis 19-jährige junge Erwachsene kurz vor den gymnasialen Endprüfungen); damit auch Fokussierung auf eher hohe Kompetenzstufen,
2. Fokussierung auf den Schultypus Gymnasium,
3. Fokussierung auf die Abschätzung der individuellen „Allgemeinen Studierfähigkeit" (gymnasiales Bildungsziel) kurz vor einem möglichen Übertritt an die (schweizerischen) universitären Hochschulen,
4. Fokussierung und Passung der Kompetenzmodellierung auf das in den ersten Semestern vorausgesetzte, im Teilprojekt A ermittelte Wissen und Können,
5. Fokussierung und Passung der Kompetenzmodellierung auf die im Rahmen des Projekts EVAMAR II möglichen Aufgabenformen bei der Leistungsmessung im Projektteil C.

Für alle drei Testfächer entstanden auf diese Weise Kompetenzfelder, welche zum im Projekt A ermittelten Wissen und Können bestmöglich passen. Im Projektteil C (siehe Berichtsteil IV) wurden dann zu diesen Feldern

Aufgaben konstruiert, welche sich auf Fachinhalte aus Projekt A beziehen. Alle Testaufgaben können also sowohl in der gewählten Kompetenzmodellierung als auch in den Lehrmaterialien an Universitäten verortet werden. Diese konkrete und konsequente „Schnittstellen-Verknüpfung“ der Tests bzw. der Leistungsmessung im Projekt EVAMAR II wurde noch nie in einer derartigen inhaltlichen Nähe vorgenommen, sie ist also neuartig.

2 Der Kompetenzraster in Erstsprache

2.1 Grundlagen

Im Vergleich zu den anderen im Projekt EVAMAR II zu modellierenden Kompetenzbereichen weist der Bereich Erstsprache die folgenden Besonderheiten auf:

- Erstsprache wird (inklusive des Fremdwortschatzes aus den Herkunftssprachen wie Latein) von der in Teilprojekt A1 generierten Datenbank (siehe Abschnitt 2.4 in Kapitel II) in breiter Weise aus allen analysierten Studienfächern bedient. Konsequenterweise genügt der darauf basierende Kompetenzraster *nicht* den Anforderungen des Projekts EVAMAR II, soweit es sich um ein stark curricularbezogenes Modell mit einer strikten auf ein Germanistikstudium ausgerichteten Orientierung handelt, das wissenschaftspropädeutische Fachinhalte des Erstsprache-Unterrichts wie Epochen- und Literaturwissen, Analysetechniken u. ä. betont. Vielmehr müssen die in Erstsprache erworbenen Fertigkeiten der allgemeinen Sprachkompetenzen (Rezeption, Produktion, Sprachreflexion) im Mittelpunkt stehen.
- Sowohl die bisherige Diskussion in der einschlägigen Literatur (beispielsweise Ehlich & Steets, 2003; Feilke & Steinhoff, 2003 und Kruse, 2007 zur Schreibkompetenz; Überblick und annotierte Bibliografie in Mugabushaka, 2005) wie auch unsere Untersuchungen in den Teilprojekten A zeigen übereinstimmend, dass ausreichende kommunikative Kompetenzen in Erstsprache im Sinne linguistischer, soziolinguistischer und pragmatischer Kompetenzen[8] unabdingbar sind für die erfolgreiche Bewältigung beinahe jeden Hochschulstudiums (siehe Kapitel II) und als „germanistisch fundierte Schlüsselqualifikationen“ (Mugabushaka, 2005, S. 62) im Hochschulbereich fortlaufend zu optimieren sind. Dazu gehört auch ein breiter und auf wissenschaftlichen Sprachgebrauch, insbesondere auf die Textsorte „universitäre Fachtexte“ (beispielsweise Ehlich, 2003; Gläser, 1990, zitiert nach Thim-Mabrey, 2005) ausgerichteter Wortschatz. Im Vergleich dazu tritt die Verfügbarkeit von literaturspezifischem Wissen in den Hintergrund. Diese Kenntnisse aus dem literarischen Kernbereich des Deutschunterrichts sind zwar durchaus im Hinblick auf verschiedene Kontexte nützlich, aber wie Teilprojekt A zeigte, selbst für das Studium beispielsweise der Germanistik nicht unabdingbar.
- Der zu entwickelnde Kompetenzraster sollte anwendbar sein für alle drei Schweizer Erstsprachen Deutsch, Französisch und Italienisch.

Die theoretische Verortung des zu entwickelnden Kompetenzrasters für Erstsprache hatte in Anlehnung – bzw. allenfalls auch in Abgrenzung – zu den bisherigen neueren Sprachuntersuchungen und -betrachtungen des internationalen, europäischen, nationalen oder auch regionalen Raumes zu erfolgen. Nennenswert sind hier insbesondere das Forschungsprojekt, in welchem der Zürcher Textanalyseraster entwickelt wurde (Sieber, 1994), das Forschungsprojekt Deutsch-Englisch-Schülerleistungen-International (DESI) (DESI-Konsortium, 2006;

---

[8] Die sprachbasierten kommunikativen Kompetenzen werden unterteilt in:

1. Linguistische Kompetenzen, welche das Verfügen über Sprachwissen (Wortschatz, Semantik, Orthografie, Aussprache etc.) und Sprachregeln (Grammatik, Interpunktion etc.) meinen,
2. Soziolinguistische Kompetenzen, bei welchen es um die Beachtung der soziokulturellen Bedingungen der Sprachverwendung geht (Höflichkeitsformen, Redewendungen etc.), und
3. Pragmatische Kompetenzen, welche den funktionalen Gebrauch von Sprachfunktionen regeln und zu welchen beispielsweise korrekter Sprecherwechsel, Themenentfaltung, Kohärenz und Textsorten-Kenntnisse gerechnet werden (Europarat, 2001, S. 25; Klieme, 2003, S. 148).

Klieme, 2006; Beck & Klieme, 2007; Willenberg, 2007; DESI-Konsortium, 2008), das Programme for International Student Assessment (PISA) (Haider & Reiter, 2001, 2004; Prenzel, 2004; Schiefele et al., 2004), der Gemeinsame Europäische Referenzrahmen für Sprachen (GER) (Europararat, 2001; Schneider & North, 1999), in der Schweiz für die aktuelle (Grund-)Schulreform insbesondere die Interkantonale Vereinbarung über die Harmonisierung der obligatorischen Schule (HarmoS) (EDK, 2004a, 2004b, 2005, 2007, 2008; Maradan & Mangold, 2005; Sieber, 2006a, 2006b, 2006c) und abschliessend (aber nicht erschöpfend) die Zentralen Lernstandserhebungen in der Jahrgangsstufe 9 im Jahr 2004 in Nordrhein-Westfalen (NRW) (Leutner, Wirth & Fleischer, 2005; Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2005a, 2005b, 2007, 2008). Diese in den letzten Jahren durchgeführten empirischen Studien sind in Bezug auf Zielsetzung und Struktur (z. B. Altersstufen und Textsorten) wenig homogen und weisen kaum kongruente Kompetenzbereiche oder unterschiedlich festgelegte Niveaustufen auf. Zudem basieren diese Evaluationsstudien meist auf ziemlich unterschiedlichen Konstrukten von Lesekompetenz (Streblow, 2004, S. 275). Insbesondere der bis anhin seltene, wenn nicht gar neue Ansatz von EVAMAR II, in den Leistungserhebungen solches Wissen und Können zu testen, das sich unmittelbar aus der Sicht der Hochschulen als studiumsrelevant ergibt, erschwerte den Anschluss an diese Vielfalt bisheriger Arbeiten noch zusätzlich.

2.2 Bestehende Kompetenzraster und -modelle für Erstsprache

2.2.1 Zürcher Textanalyseraster

2.2.1.1 Kurzbeschreibung

Im Rahmen des Nationalfondsprojekts „Muttersprachliche Fähigkeiten von Maturanden und Studienanfängern in der Deutschschweiz“ (Sieber, 1994) wurden die Erstsprachleistungen von Maturandinnen und Maturanden sowie Studienanfängerinnen und Studienanfängern gemäss unserer Kenntnisse erstmals in der Schweiz gleichzeitig untersucht. Es wurde sowohl der Ist-Zustand erfasst als auch der Soll-Zustand, letzterer durch Erhebung der Erwartungen von Lehrpersonen an Mittel- und Hochschulen mittels Umfragen und halbstrukturierten Interviews. Die Datengrundlage für die Analyse der Sprachfähigkeiten bildeten 800 Aufsätze von Maturanden und Maturandinnen der Abschlussjahre 1988-90 aus zehn Deutschschweizer Mittelschulen, wobei vorwiegend Schülertexte aus dem Unterricht (Deutsch, Physik, Biologie) sowie einige Matura-Prüfungsaufsätze zur Verfügung standen. Dazu kam ein kleiner Korpus an Texten von Studienanfängerinnen und -anfängern der Germanistik der Universität Zürich (S. 191-200). Sieber und sein Team entwickelten einen differenzierten Textproduktions-Raster zur Analyse und Erfassung der Aufsätze, den so genannten Zürcher Textanalyseraster (Sieber, 1994, S. 149-186). Den theoretischen Hintergrund für dieses Instrument, das Aussagen über Qualitäten und Mängel von Texten ermöglicht, bildet eine linguistische Texttheorie (Nussbaumer, 1991, 1993, zitiert in Sieber, 1994, S. 143-148), welche den Text als integrale, kohärente Einheit versteht, die von den Rezipienten durch die Aktualisierung von Sprachwissen, Weltwissen und Handlungswissen rekonstruiert werden muss. „Im Sinne einer heuristischen Folie“ (S. 43) nutzen Sieber et al. das Schichtenmodell der sprachlichen Kompetenz des Sprachwissenschaftlers Glinz (1987, S. 206-236, zitiert nach Sieber, 1994, S. 41-43). Dieses basiert auf sieben Kompetenzbereichen mit abnehmender Relevanz (siehe Abbildung III.1).

Abb. III.1: Schichtenmodell der Sprachfähigkeiten von Glinz (1987, zitiert nach Sieber, 1994, S. 41)

I	Zentralbereich der Personen-Konstitution, <Aufbau von Identität>
II	Muster für Textaufbau und Strategien der Sprachverwendung
III	Semantischer Bereich: Bedeutungen und Bedeutungsstrukturen
IV	Lexisch-morphosyntaktischer Bereich: Wörter und grammatische Formalstrukturen
V	Ganze Stimmführungsgestalten, Klangliches oberhalb der Wortlautungen
VI	Lautungssystem, Phono-Morphologie, <Aussprache>
VII	Schreibungssystem, Orthografie und Interpunktion

Sieber et al. (1994) interpretieren die Ergebnisse ihrer Sprachfähigkeitsuntersuchungen (S. 308 ff.) vor dem Hintergrund dieses Glinzschen Kompetenzmodells, nehmen allerdings eine Modifikation vor: Sie ziehen die Glinzschen Teilbereiche IV und VII zu dem neuen Bereich „Wörter und grammatische Formalstrukturen/Schreibung“ sowie die ursprünglichen Bereiche V und VI zu dem Gebiet „Stimmführung und Lautung“ zusammen. Anhand dieser fünf Bereiche werden die schriftlichen und mündlichen Fähigkeiten der Maturandinnen und Maturanden bzw. der Studierenden beurteilt.

2.2.1.2 Einschätzung des Zürcher Textanalyserasters für die Verwendung in EVAMAR II

Die Beschreibungen des Zürcher Textanalyserasters (Sieber, 1994) sind insofern mit den Vorhaben im Projekt EVAMAR II vergleichbar, als das Datenmaterial sich grösstenteils ebenfalls auf die auf den Übertritt an die Universität ausgerichtete gymnasiale Abschlussstufe bezieht.

Allerdings basieren die Ergebnisse des Sieberschen Projekts für den Teilbereich der Kompetenzen im schriftlichen Bereich auf Schriftstücken, welche aufgrund ihrer Länge auf wichtige Merkmale der Textproduktion, wie Kohärenz, Disposition und Leserführung, untersucht werden können. Wie die weiteren Ausführungen zeigen werden, musste aber im Verlauf der EVAMAR II-Arbeiten im Rahmen der Tests auf die aufwändige Textproduktion von langen Schriftstücken wie Briefen, Erörterungen, Analysen zugunsten kleinerer offener Formate verzichtet werden. Der zweite von den Lehrpersonen und Experten eingeschätzte Teilbereich der mündlichen Ausdrucksfähigkeit war für die Ausrichtung von EVAMAR II ebenfalls wenig relevant. Damit erwies sich eine mögliche Verortung des EVAMAR II-Erstsprache-Kompetenzrasters im Umfeld des Zürcher Textanalyserasters als hinfällig.

2.2.2 DESI

2.2.2.1 Kurzbeschreibung

Das bislang grösste deutschdidaktische Forschungsprojekt Deutschlands „Deutsch-Englisch-Schülerleistungen-International“ (DESI) wurde in der PISA-Nachfolge von der Deutschen Kultusministerkonferenz (KMK) für die Jahre 2001 bis 2006 in Auftrag gegeben, um den Sprachunterricht in allen Aspekten repräsentativ zu testen. Zu diesem Zweck wurden unter der Projektleitung des Deutschen Instituts für Internationale Pädagogische Forschung (DIPF) rund 10 640 Jugendliche aller Schularten aus 219 Schulen aller Bundesländer am Anfang und Ende der neunten Klasse getestet, um für alle Teilbereiche des Deutschunterrichtes wissenschaftlich fundierte und empirisch bestätigte Kompetenzmodelle zu gewinnen. Hinzu kamen auch kompetenzmodellorientierte Video-Beobachtungen von Unterrichtsstunden (Englisch). Die Tests basieren auf Analysen der curricularen Vorgaben der Bundesländer, beispielsweise nach Textsorten im Schreiben oder im Bereich Sprachbewusstheit nach Wortarten.

Der DESI-Test Deutsch beruht auf sechs Kompetenzbereichen mit drei bis fünf unterschiedenen Kompetenzniveaus[9] (A bis E). In den Bereichen Schreibkompetenz, Wortschatz, Rechtschreiben, Argumentation werden drei, im Kompetenzbereich Lesen vier und im Kompetenzbereich Sprachbewusstheit fünf Kompetenzniveaus unterschieden. Den 2006 veröffentlichten zentralen Befunden der Studie (DESI-Konsortium, 2006; Klieme, 2006; Beck & Klieme, 2007; Willenberg, 2007) lassen sich die Kann-Beschreibungen zu drei Kompetenzbereichen entnehmen. Sie sind in Abbildung III.2 zusammengetragen.

Abb. III.2: Eigene Darstellung von drei der sechs DESI-Kompetenzbereiche für Deutsch (DESI-Konsortium, 2006) mit zugehörigen Niveaus A-E (ohne Argumentation, Wortschatz, Rechtschreiben)

Kompetenz-niveau	Sprachbewusstheit (Tabelle 3, ergänzt für E mit Beschreibungen[10] im Text, S. 9)	Lesen (Tabelle 2, S. 5)	Schreibkompetenz (Beschreibung S. 7)	
			Semantisch-pragmatische Teilkompetenz	Sprachsystematische Teilkompetenz
E	**Aktive Anwendung deklaratorischen Wissens:** Explizites (deklaratorisches) Wissen über komplexe Phänomene wie die indirekte Rede und die Formen des Konjunktivs aktiv nutzen.	*(Kein Niveau E)*	*(Kein Niveau E)*	*(Kein Niveau E)*
D	**Entfaltetes grammatisches und stilistisches Monitoring**: Komplexe Stilfehler erkennen und beseitigen, Mehrdeutigkeiten auflösen.	**Auswerten mentaler Modelle:** Hauptfiguren und ihre Relationen, Ort, Zeit sowie ein zentrales Motiv zusammenfügen.	*(Kein Niveau D)*	*(Kein Niveau D)*
C	**Entfaltetes grammatisches Monitoring:** Sicherheit im Erkennen und Korrigieren auch schwieriger grammatischer Phänomene. Einfache Zuordnung grammatischer Begriffe zu Phänomenen.	**Verknüpfende Lektüre:** Aktivieren von allgemeinem Wissen bzw. Textwissen; Verknüpfen zweier auseinanderliegender Stellen, meistens um Motive oder Kausalitäten zu klären.	Stilistisch anspruchsvoll, abwechslungsreich in der Wortwahl und konsequent logisch aufgebaut sind die Briefe auf Niveau C.	(Keine Angaben)
B	**Einfache grammatische und stilistische Sprachbewusstheit:** Stil und Kohärenz in einfachen Kontexten und Inhaltsbereichen herstellen.	**Fokussierte Lektüre:** Inferenzen bilden, besonders bei Lücken zwischen zwei Sätzen; genaues, fokussierendes Lesen an einer semantisch oder logisch schwierigen Stelle; Koordination von Zahlen in den Texten.	Auf diesem Niveau ist eine ungestörte schriftliche Kommunikation zwischen Absender und Empfänger des Schreibens möglich. Einen Brief mit eigenem z. T. noch begrenztem Wortmaterial im Wesentlichen schlüssig aufbauen, sodass die zentralen Elemente eines Briefes zumindest in Ansätzen zu erkennen sind.	Orthographische Konventionen werden weitgehend eingehalten und angemessen komplexe Satzkonstruktionen verwendet.

[9] Die Unterteilung der kontinuierlichen Kompetenzskala in Abschnitte erfolgt bei DESI durch den Begriff Kompetenz*niveaus*. DESI verzichtet bewusst auf die beispielsweise bei PISA übliche Übersetzung des englischen Begriffes *proficiency level* durch *Stufe*, weil dieser fälschlicherweise ein Stufenmodell im Piagetschen Sinn konnotiere (Hartig, 2007, S. 86).

[10] Die später veröffentlichten Beschreibungen zur Sprachbewusstheit variieren zum Teil (Beck & Klieme, 2007, S. 154-157); da diese uns in der Modellierungsphase noch nicht vorlagen, verzichten wir auf deren Wiedergabe (Beispiel Niveau E (S. 156): „Aktive Anwendung deklaratorischen Wissens in Form von bewusster Kommentierung, prozedurales Wissen auch in hochkomplexen Zusammenhängen“).

	Sprachbewusstheit (Tabelle 3, ergänzt für E mit Beschreibungen[10] im Text, S. 9)	Lesen (Tabelle 2, S. 5)	Schreibkompetenz (Beschreibung S. 7)	
A	**Einfache grammatische Sprachbewusstheit**: Eindeutige grammatische Fehler erkennen und z. T. korrigieren.	**Identifizierende Lektüre:** Erkennen der sinntragenden Einheiten in einem Satz oder einem Abschnitt.	Entspricht einer Textproduktion, die eher umgangssprachlichen Sprachnormen folgt, logisch fehlerhaft aufgebaut ist und wesentliche Textelemente eines Briefes nicht enthält.	Fehler in Orthografie, Zeichensetzung und Satzkonstruktion, die sich kommunikationsbelastend auswirken oder die eine zielgerichtete Informationsübertragung aufgrund sprachsystematischer Mängel unmöglich machen.

2.2.2.2 Einschätzung des Kompetenzmodells von DESI für die Verwendung in EVAMAR II

Das in DESI verwendete und erfolgreich validierte Kompetenzmodell wurde für Tests der neunten Jahrgangsstufe entwickelt und orientiert sich an Curricula-Inhalten der 16 deutschen Bundesländer. Diese Rahmenbedingungen hätten für EVAMAR II übertragen werden müssen auf die 12. Jahrgangsstufe bzw. auf universitäre Anforderungen und wissenschaftliche Textsorten. Eine Ausdifferenzierung und Erweiterung des DESI-Modells unter Verwendung der obersten Skalenbereiche wäre unumgänglich gewesen.

Die Entwicklung des DESI-Modells für den Kompetenzbereich „Textproduktion“ anhand der Textsorte „Brief“ (zur einen Hälfte wurden so genannte „Sachbriefe“ und zur anderen Hälfte persönliche Briefe eingesetzt) war für unsere Zwecke nachteilig. Damit war ein eventueller Textsorten-Vergleich für EVAMAR II-Anliegen von vornherein ungeeignet, allenfalls höchstens in Ansätzen möglich, da die eingeforderten Textmuster auf themengebundene Aufsätze und Erörterungen hätten erweitert werden müssen, sowie allenfalls auf (Pro-)Seminar-Arbeiten.

Für den interessant erscheinenden Kompetenzbereich Argumentation erwies es sich bei erweiterter Betrachtung für EVAMAR II als von Nachteil, dass diese Fertigkeit bei DESI als Analysekompetenz bereits vorhandener Argumentationsformen und -intentionen erfasst wurde, es sich hier somit ausschliesslich um rezeptive Wiedererkennungsleistungen handelt.

Als weitere Erschwernis lagen in der Entscheidungsphase des Projektes EVAMAR II die Anschlusspublikationen der inzwischen implementierten Modelle der Schülerleistungsstudie noch nicht vor, welche die theoretischen Herleitungen der Teilkompetenz-Modelle und deren Umsetzung in der empirischen Fundierung ausführlich interpretieren (Beck & Klieme, 2007; Willenberg, 2007; DESI-Konsortium, 2008).

So erwies die damalige Prüfung, die im Zusammenhang mit der Entwicklung eines für die Schnittstelle Gymnasium – Hochschule geeigneten Kompetenzrasters Erstsprache erfolgte, das kurz vorgestellte Modell von DESI als nur mittelmässig geeignet, um Verortungsfeld für die spezifische Ausrichtung von EVAMAR II zu werden.

2.2.3 PISA

2.2.3.1 Kurzbeschreibung

Die internationale PISA-Studie der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) orientiert sich an der angelsächsischen Literacy-Konzeption. Diese versteht Kompetenzen als basale Kulturwerkzeuge und weist insbesondere dem Lesen als Schlüsselqualifikation eine grosse funktionale Rolle als Voraussetzung für den Umgang mit allen Medien zu (Artelt, Stanat, Schneider, Schiefele & Lehmann, 2004, S. 141).

Die Leistungstests enthalten eine umfassende Bandbreite von Lesesituationen, die sich an praktischem Alltagshandeln orientieren und neben kontinuierlichen Texten auch so genannte diskontinuierliche Texte (Verknüp-

fung mit Tabellen, Grafiken) anbieten. 70% der Aufgaben beziehen sich auf textimmanente Verstehensleistungen; 30% gehen darüber hinaus, d. h. erfordern auch ein Reflektieren und Bewerten. 45% der Items sind als offene Formate konstruiert (Artelt et al., 2004, S. 142).

Im Unterschied zum breit gefassten Sprachmodell von DESI bildet PISA für die Sprachen ausschliesslich den Teilbereich ‚Lesen' ab. Im theoretischen Konzeptionsrahmen (OECD, 1999/2000, 2001, zitiert nach Artelt & Schlagmüller, 2004, S. 169) waren ursprünglich fünf Aspekte des Lesens definiert. Mit den PISA-Leseaufgaben sollten die Fähigkeiten abgebildet werden, erstens Informationen zu ermitteln, zweitens ein allgemeines Verständnis des Textes und drittens eine textbezogene Interpretation zu entwickeln, viertens über den Inhalt des Textes und fünftens über seine Form reflektieren zu können (Artelt et al., 2004, S. 143; Artelt & Schlagmüller, 2004, S. 173). Bei der Testkonzeption selbst lag noch kein zugrunde liegendes Kompetenzmodell vor (Artelt et al., 2004, S. 151).

Durch die in PISA 2000 erhobenen empirischen Daten konnten später nur drei der fünf Teilskalen von Teilkompetenzen bestätigt werden: erstens Informationen ermitteln, zweitens textbezogenes Interpretieren und drittens Reflektieren und Bewerten (Artelt & Schlagmüller, 2004, S. 175).

Basierend auf der Vorgabe, gleich breite Niveaubereiche zu bilden (Artelt et al., 2004, S. 144), wurden zu diesen Lese-Teilkompetenzen fünf Kompetenzstufen formuliert. Zu jeder Stufe liegt eine Darstellung der charakteristischen Anforderungsmerkmale in Kurzbeschreibungen sowie illustrierenden Beispielaufgaben vor (siehe Abbildung III.3).

Abb. III.3: Das PISA-Lesekompetenzmodell mit fünf Kompetenzstufen. Eigene Darstellung der drei Lese-Teilkompetenzen (nach Artelt et al., 2004, S. 144-145).

PISA 2000 Lese-Kompetenz	Informationen ermitteln	Textbezogenes Interpretieren	Reflektieren und Bewerten
Stufe 5: *Flexible Nutzung unvertrauter komplexer Texte*	Expertenleserinnen und -leser; verstehen auch komplexe unvertraute und lange Texte vollständig und detailliert. Auch Elemente, die ausserhalb des Hauptteils des Textes liegen und die in starkem Widerspruch zu den eigenen Erwartungen liegen.	Die Bedeutung feiner sprachlicher Nuancen wird angemessen interpretiert.	Expertenleserinnen und -leser; können lange, unvertraute, komplexe Texte in ihr Vorwissen aus verschiedenen Bereichen einbetten und auf dieser Grundlage kritisch bewerten.
Stufe 4: *Detailliertes Verständnis komplexer Texte*	Potenzielle Hürden eines relativ unvertrauten Textes (wie Mehrdeutigkeiten, Sprachnuancen oder den eigenen Erwartungen widersprechende Elemente) können weitgehend bewältigt werden.	Ein genaues Verständnis komplexer, relativ langer und im Hinblick auf Form und Inhalt relativ unvertrauter Textewird erreicht.	Relativ lange, komplexe Texte können unter Rückgriff auf externes Wissen beurteilt werden.
Stufe 3: *Integration von Textelementen und Schlussfolgerungen*	Mit relativ auffällig konkurrierenden Informationen, die von relevanten Elementen ablenken könnten, kann umgegangen werden. Verschiedene Teile des Textes können integriert werden, auch wenn die einzubeziehende Information wenig offensichtlich ist und ihre Bedeutung teilweise indirekt erschlossen werden muss.	Mittleres Anspruchsniveau: Ein genaues Verständnis von Texten mittleren Komplexitätsgrades entwickeln können.	Das Gelesene kann auf der Grundlage von spezifischem Wissen beurteilt werden.

PISA 2000 Lese-Kompetenz	Informationen ermitteln	Textbezogenes Interpretieren	Reflektieren und Bewerten
Stufe 2: *Herstellen einfacher Verknüpfungen*	Die Bedeutung einzelner Elemente kann durch simple Schlussfolgerung erschlossen werden.	Der Hauptgedanke eines relativ vertrauten Textes kann identifiziert und ein breites Textverständnis entwickelt werden.	Die gelesenen Informationen können unter Bezugnahme auf persönliche Erfahrungen und Einstellungen beurteilt werden.
Stufe 1: *Oberflächliches Verständnis einfacher Texte*	Die notwendige Information im Text muss deutlich erkennbar sein und der Text darf nur wenige konkurrierende Elemente enthalten, die von der relevanten Information ablenken könnten.	Elementare Lesefähigkeiten: Mit einfachen Texten umgehen können, die in Form und Inhalt vertraut sind.	Es können nur relativ offensichtliche Verbindungen zwischen dem Gelesenen und allgemein bekanntem Alltagswissen hergestellt werden.

Interessant war im Hinblick auf Projektteil C von EVAMAR II die Feststellung, dass die schwierigkeitsgenerierenden Merkmale von Aufgaben der verschiedenen Kompetenzstufen im Vorfeld der Testung kaum bestimmt werden konnten (Artelt et al., 2004, S. 153).

Quer zu den beschriebenen Lese-Aspekten steht die Untersuchung der Einflüsse des Textgenres, der Art des Textes und der Textformate auf die erfassten Lesekompetenzen. Bei kontinuierlichen Texten ist u. a. die Frage der Kohärenz, der Propositionsdichte und des Gebrauchs von Metaphern und Analogien von Bedeutung; bei diskontinuierlichen Texten treten durch die Bildverarbeitungsprozesse andere schwierigkeitsbestimmende Merkmale auf. Es liess sich bei PISA 2000 empirisch nachweisen, dass „die Unterscheidung der Lesekompetenz in spezifische Kompetenzen beim Umgang mit kontinuierlich und nichtkontinuierlich geschriebenen Texten von den Zahlenwerten her vertretbar ist“ (Artelt & Schlagmüller, 2004, S. 176, Hervorh. Autoren).

Als zusätzlicher Beleg für die Konstruktvalidität dieser Unterscheidung gilt das differenzielle Ergebnismuster in den Leistungen von Mädchen und Jungen bei PISA: Die Mädchen erzielen beim Lesen nichtkontinuierlicher Texte einen signifikant kleineren Vorsprung gegenüber den Knaben als beim Lesen von kontinuierlichen Texten (Artelt & Schlagmüller, 2004, S. 176-177, S. 183-185).

Auch der Umgang mit literarischen Texten kann als eine eigenständige Kompetenzdimension gesehen werden (Artelt & Schlagmüller, 2004, S. 178).

Versucht man nun, von den Ergebnis-Tabellen und Schlussfolgerungen abstrahierend, sich für eine Kompetenzraster-Entwicklung das dahinter stehende Modell vorzustellen, so würde man unter Verwendung der empirisch bestätigten PISA-Lesen-Teilkompetenzen und der fünf verwendeten PISA-Kompetenzstufen und der verschiedenen analysierten Textsorten zu folgender Darstellungsmatrix eines Lesemodells gelangen (Abbildung III.4).

Abb. III.4: Eigene Darstellung: Das PISA-Lesemodell unter Einbezug der verwendeten Textsorten

PISA 2000 Lese-Kompetenz	Diskontinuierliche Texte	Kontinuierliche nichtliterarische Texte	Literarische Texte
Informationen ermitteln	*Stufe 5*	*Stufe 5*	*Stufe 5*
	Stufe 4	*Stufe 4*	*Stufe 4*
	Stufe 3	*Stufe 3*	*Stufe 3*
	Stufe 2	*Stufe 2*	*Stufe 2*
	Stufe 1	*Stufe 1*	*Stufe 1*
Textbezogenes Interpretieren	*Stufe 5*	*Stufe 5*	*Stufe 5*
	Stufe 4	*Stufe 4*	*Stufe 4*
	Stufe 3	*Stufe 3*	*Stufe 3*
	Stufe 2	*Stufe 2*	*Stufe 2*
	Stufe 1	*Stufe 1*	*Stufe 1*
Reflektieren und Bewerten	*Stufe 5*	*Stufe 5*	*Stufe 5*
	Stufe 4	*Stufe 4*	*Stufe 4*
	Stufe 3	*Stufe 3*	*Stufe 3*
	Stufe 2	*Stufe 2*	*Stufe 2*
	Stufe 1	*Stufe 1*	*Stufe 1*

2.2.3.2 Einschätzung des PISA-Modells für die Verwendung in EVAMAR II

Bezogen auf die Anforderungen von EVAMAR II wäre das Pisa-Modell, konzentriert auf die obersten Kompetenzstufen und erweitert um eine zusätzliche und neu zu definierende Stufe 6, adaptierbar.

Es könnte damit zwar nur der Teilbereich „Lesen“ abgedeckt werden. Die Subskalen „Informationen ermitteln – Textbezogenes Interpretieren – Reflektieren und Bewerten“ könnten als bereits empirisch validierte Teilkompetenzen von „Lesen“ übernommen werden, wobei die Adaption an komplexere mentale Prozessstrukturen geprüft werden muss. Die Einteilung in diskontinuierliche und kontinuierliche bzw. literarische Texte ist ebenfalls geeignet, wobei die kontinuierlichen nicht-literarischen Texte überwiegen und nur in einzelnen Studienfächern vereinzelt auch andere Textsorten vorkommen werden (z. B. literarische bei Germanistik, diskontinuierliche bei Mathematik). Insgesamt müsste man das Profil der Tests für das 9. Schuljahr übertragen auf die Bedingungen des 12./13. Schuljahrs in Gymnasien und auf universitäre Anforderungen.

2.2.4 Der Gemeinsame Europäische Referenzrahmen (GER)

2.2.4.1 Kurzbeschreibung

Der vom Europarat 2001 verabschiedete Gemeinsame Europäische Referenzrahmen für Sprachen (GER) bietet einen umfassenden, transparenten und kohärenten Referenzrahmen für das Lernen und Lehren von Sprachen. Er beruht auf einer seit 1971 geführten Diskussion unter Sprachexperten und -expertinnen aus 40 Ländern und bildet in ganz Europa eine gemeinsame Basis, welche die sprachenpolitischen Ziele des Europarates in konkreten Werkzeugen umsetzt. Der Referenzrahmen wird zielorientiert eingesetzt für die Entwicklung von Sprach-Lehrplänen, curricularen Richtlinien, Prüfungen und Lehrwerken. Die Kompetenzniveaus sind definiert für lebenslanges Lernen (Europarat, 2001, S. 14).

Der Ansatz des GER ist ein handlungsorientierter: Die Sprachverwendenden werden als sozial Handelnde verstanden; demzufolge erfolgen die Kann-Beschreibungen ihrer Kenntnisse und Fähigkeiten innerhalb gegebener Handlungs- und Sprachverwendungsfelder (Europarat, 2001, S. 21, 26). Hierbei fokussiert der GER die wichtigsten Domänen: Es geht um Sprachverwendung im Bildungswesen, im beruflichen, privaten und öffentlichen Bereich (S. 26). Die Referenzskalen selbst sollen jedoch kontextfrei sein, d. h. sie sind für verschiedene spezifische Kontexte generalisierbar. Nicht zuletzt dank seiner Einfachheit ist der GER gut handhabbar und die elektronischen und gedruckten Mittel zur Selbsteinschätzung, Fremdbeurteilung und Testung ermöglichen einen vielfältigen Einsatz im schülergerechten Sprachunterricht und in der Lehrplan- und Testentwicklung.

Der GER wird dem Anspruch gerecht, einen Rahmen für die Bereitstellung objektiver Kriterien zur Beschreibung von Sprachkompetenzen auf breiter Ebene zu bieten. Dazu gehören erstens die allgemeinen Kompetenzen der Sprachverwendenden, ihr deklaratives Wissen (savoir), ihre Fertigkeiten und das prozedurale Wissen (savoir-faire), die persönlichkeitsbezogene Kompetenz und die Einstellungen (savoir-être), sowie die Lernfähigkeit (savoir-apprendre). Zweitens werden auch als zur kommunikativen Sprachkompetenz gehörig betrachtet die linguistischen Kompetenzen (lexikalische, phonologische und syntaktische Kenntnisse und Fertigkeiten), die soziolinguistischen Kompetenzen (Bedingungen der gesellschaftlichen Konventionen kennen) sowie die pragmatischen Kompetenzen; die Letzteren regeln den funktionalen Gebrauch sprachlicher Ressourcen und meinen beispielsweise die Identifikation von Textsorten und -typen, den richtigen Gebrauch von Ironie oder das Verstehen und Herstellen von Kohärenz und Kohäsion[11]. Die konkrete kommunikative Sprachaktivität (language activity) wird beschrieben in den Bereichen der Rezeption, der Produktion, der Interaktion und der Sprachmittlung, womit hauptsächlich Dolmetschen oder Zusammenfassen gemeint ist (Europarat, 2001, S. 22-26).

Der relevante Kompetenzraum für Sprachlernende wird im GER durch sechs breit angelegte Niveau-Stufen von A1 bis C2 abgedeckt, welche der klassischen Lernbereichsaufteilung von Grund-, Mittel- und Oberstufe zugeordnet werden können. A1 ist die vermutlich niedrigste identifizierbare Stufe einer unabhängigen Sprachkompetenz. Darunter lassen sich jedoch schon einfache, allgemeine Aufgaben formulieren, welche vorausgesetzt sind (wie „einfache Grussformeln verwenden“, „unkomplizierte Formulare mit Namen etc. ausfüllen“) (Europarat, 2001, S. 40). Die Elementare Sprachverwendung A ist die Anfängerstufe im Sprachenlernen und ist in zwei Schritte (A1 Breakthrough und A2 Waystage) unterteilt. Danach folgt die Ebene der Selbstständigen Sprachverwendung B, welche auch wieder in zwei Stufen gegliedert ist (B1 Treshold und B2 Vantage), die letztere meint bereits eine „angemessene Reaktion auf Situationen, denen man normalerweise begegnet“ (Europarat, 2001, S. 33). Die dritte Ebene der Kompetenten Sprachverwendung C wird unterteilt in das fortgeschrittenere Kompetenzniveau C1 (Effective Operational Proficiency), das die Bewältigung komplexer kommunikativer Aufgaben in Beruf und Studium ermöglicht und die höchste Stufe C2 Mastery, welche das oberste Prüfungsziel innerhalb des Systems der Association of Language Testers in Europe (ALTE) bildet (Europarat, 2001, S. 40-45).

Zur Beschreibung der konkreten Kompetenzen innerhalb der sechs unterschiedenen Niveaustufen A1 bis C2 wurden 1993 bis 1996 innerhalb des Schweizerischen Nationalfondsprojekts 33 „Die Wirksamkeit unserer Bildungssysteme“ (Schneider & North, 1999) so genannte positive Kann-Beschreibungen und Beispiel-Deskriptoren entwickelt und mit der Rasch-Methode skaliert (Europarat, 2001, S. 210-217). Die Untersuchungsbasis bildeten insgesamt rund 2800 Lernenden in 500 Klassen der Sekundarstufe I und II, in der Berufs- und Erwachsenenbildung. Dieses Schweizer Projekt von Schneider und North zur Entwicklung von Deskriptoren konnte empirisch weitere Zwischenstufen (oder so genannte Plus-Niveaus: A2+/B1+/B2+) nachweisen, welche eine kohärente Skala von neun Niveau-Stufen (Europarat, 2001, S. 40, S. 210-213) respektive zehn Niveaus (Schneider & North, 1999, S. 17) nahe legten; in der Praxis wird aber aus politischen und historischen Gründen nicht auf die so genannte Schweizer Skala, sondern meist auf oben beschriebene sechs Anforderungsstufen referiert (Schneider & North, 1999, S. 16; Europarat, 2001, S. 33).

[11] *Kohärenz* bezeichnet textlinguistisch die Tiefenstruktur eines Textes, *Kohäsion* die textuelle Oberfläche – Textkompetenz meint hier die Fähigkeit einer Person, „vom Verstehen der sprachlichen Bestandteile der Textoberfläche zum Sinnverständnis des Textes vorzudringen“ (Hornung, 2005, S. 129).

Die Globalskala (siehe Abbildung III.5) beinhaltet allgemeine Kann-Beschreibungen für alle sechs Niveaus, im globalen Selbstbeurteilungs-Raster (S. 36) finden wir diese geordnet nach den Sprachverwendungsbereichen Verstehen, Sprechen, Schreiben.

Abb. III.5: Gemeinsame Referenzniveaus: Globalskala (Auszug Niveau C) des GER (Europarat, 2001, S. 35)

Kompetente Sprachverwendung	C2	Kann praktisch alles, was er/sie liest oder hört, mühelos verstehen. Kann Informationen aus verschiedenen schriftlichen und mündlichen Quellen zusammenfassen und dabei Begründungen und Erklärungen in einer zusammenhängenden Darstellung wiedergeben. Kann sich spontan, sehr flüssig und genau ausdrücken und auch bei komplexeren Sachverhalten feinere Bedeutungsnuancen deutlich machen.
	C1	Kann ein breites Spektrum anspruchsvoller, längerer Texte verstehen und auch implizite Bedeutungen erfassen. Kann sich spontan und fliessend ausdrücken, ohne öfter deutlich erkennbar nach Worten suchen zu müssen. Kann die Sprache im gesellschaftlichen und beruflichen Leben oder in Ausbildung und Studium wirksam und flexibel gebrauchen. Kann sich klar, strukturiert und ausführlich zu komplexen Sachverhalten äussern und dabei verschiedene Mittel zur Textverknüpfung angemessen verwenden.

Zur differenzierteren Beurteilung werden in jedem Sprachverwendungsbereich mehrere Subskalen zur Verfügung gestellt.

Im Bereich Produktion handelt es sich im Mündlichen um den Beurteilungsraster zur mündlichen Kommunikation (Schneider & North, 1999, S. 22; Europarat, 2001, S. 37-38) und um fünf Subskalen zum zusammenhängenden, argumentierenden und vortragenden Sprechen (Europarat, 2001, S. 64-66), sowie um mehrere Subskalen zur mündlichen Interaktion wie „zusammenhängend sprechen“ oder „an Gesprächen teilnehmen“ (Europarat, 2001, S. 79-85).

Im Schriftlichen legt der Europarat (2001) Skalen vor zur schriftlichen Produktion allgemein, zum kreativen Schreiben und zum Verfassen von Berichten und Aufsätzen[12] (S. 67-68), ferner zur schriftlichen Interaktion (S. 85-87) sowie zum „Texte verarbeiten“ und zum „Notizen machen (in Vorträgen, Seminaren etc.)“ (S. 98). Als allgemeine Produktionsstrategien werden mit Planen, Kompensieren und Kontrolle & Reparaturen drei weitere Subskalen angeboten (S. 70-71). Zu Strategien der Sprachmittlung (z. B. Zusammenfassen, Paraphrasieren von Fachtexten) gibt es noch keine Skalen.

Im Bereich der Rezeption finden sich folgende fünf Beispielskalen für das Lesen: Leseverstehen allgemein, Korrespondenz lesen, zur Orientierung lesen, Informationen und Anleitungen verstehen, schriftliche Anweisungen verstehen (S. 74-76).

Für die audiovisuelle Rezeption gibt es die Skala „Fernsehsendungen und Filme verstehen“ (S. 77) und zum Hörverstehen mehrere Subskalen (S. 71-73). Zusätzlich liegt zu den Rezeptionsstrategien, mit welchen der Kontext auf der Basis relevanten Weltwissens erschlossen wird, eine Teilskala „Hinweise identifizieren“ vor (S. 78).

Im weiteren Bereich kommunikativer Sprachkompetenzen wird unter den linguistischen Kompetenzen die Skala „Spektrum sprachlicher Mittel (allgemein)“ angeboten, zur lexikalischen Kompetenz gehören die Subskalen „Wortschatzspektrum“ (S. 112) und „Wortschatzbeherrschung“ (S. 113), zur grammatischen Kompetenz die Skala „Grammatische Korrektheit“ (S. 114) und zur orthographische Kompetenz die Skala „Beherrschung der Orthografie“ (S. 117-118). Des Weiteren erscheinen die Skalen „soziolinguistische Angemessenheit“ (S. 121-122)

[12] Die Deskriptoren dieser drei schriftlichen (Sub-)Skalen wurden nicht empirisch kalibriert, sondern durch eine Kombination von Elementen anderer Skalen erstellt (S. 67). Dies betrifft z. T. auch Subskalen des Mündlichen sowie die orthographische Skala.

und als pragmatische Kompetenzen im Bereich der Diskurskompetenzen die vier Subskalen „Flexibilität“, „Sprecherwechsel“, „Themenentwicklung“, „Kohärenz und Kohäsion“ (S. 123-125) sowie „Flüssigkeit (mündlich)“ und „Genauigkeit“ (S. 129).

Im Rahmen des EU-Projektes DIALANG wurden für Lesen, Schreiben und Hören mit der finanziellen Unterstützung der Europäischen Kommission weitere Deskriptoren empirisch validiert. Man skalierte 107 Items und kalibrierte sie mit der Rasch-Methode auf der Basis finnischer Daten am damaligen Koordinationszentrum der Universität von Jyväskylä (Europarat, 2001, S. 218-231). Ein Auszug für Niveau C findet sich in Abbildung III.6.

Für 14 europäische Sprachen wurden damit (aus dem Englischen übersetzt) ein Selbsteinschätzungsteil, ein Sprachtest und beratende Rückmeldungen entwickelt, die alle Sprachlernenden im Internet kostenlos beziehen können (dialang.org).

Abb. III.6: DIALANG-Items zur Selbsteinschätzung – Auszug Niveau C (Europarat, 2001, S. 223-226)

Lesen	
C1	Ich kann eine Reihe langer, komplexer Texte genau verstehen, vorausgesetzt ich kann schwierige Abschnitte mehrfach lesen.
C1	Ich kann in unterschiedlichen langen und komplexen Texten, die das gesellschaftliche, berufliche oder akademische Leben betreffen, feinere Bedeutungsnuancen erkennen, auch was Einstellungen und Meinungen anbelangt, die zwischen den Zeilen stehen.
C1	Unter gelegentlicher Zuhilfenahme des Wörterbuches kann ich jegliche Art von Schriftverkehr verstehen.
C1	Ich kann lange, komplexe Anleitungen für neue Geräte oder neue Verfahren auch ausserhalb meines Fachgebietes genau verstehen, wenn ich schwierige Passagen mehrmals lesen kann.
C2	Ich kann nahezu jegliche Form geschriebener Sprache verstehen und interpretieren, einschliesslich abstrakter, bezüglich der Struktur komplexer oder stark umgangssprachlicher literarischer und nicht-literarischer Schriftstücke.
C2	Ich kann ein breites Spektrum langer und komplexer Texte verstehen und dabei Feinheiten bezüglich Stil und Bedeutung, welche implizit oder explizit ausgedrückt sind, vollständig erfassen.

Schreiben	
C1	Ich kann Standpunkte durch zusätzliche Argumente, Begründungen und sinnvolle Beispiele weiter ausführen und untermauern.
C1	Ich kann eine Erörterung systematisch aufbauen und dabei wesentliche Punkte angemessen hervorheben sowie unterstützende Einzelheiten mit gewisser Bedeutung anführen.
C1	Ich kann komplexe Sachverhalte klar und detailliert schriftlich darstellen.
C1	Ich kann normalerweise ohne Zuhilfenahme eines Wörterbuchs schreiben.
C1	Ich kann so gut schreiben, dass meine Ausdrucksweise nur dann überprüft werden muss, wenn es sich um ein wichtiges Schriftstück handelt.
C2	Ich kann mich schriftlich erfolgreich, angemessen und gut strukturiert ausdrücken, sodass der Leser die wesentlichen Punkte leicht erfassen kann.
C2	Ich kann klare und flüssige komplexe Berichte, Artikel oder Aufsätze verfassen, die einen Sachverhalt darstellen oder eine kritische Bewertung von Anträgen oder literarischen Werken abgeben.
C2	Ich kann immer ohne Zuhilfenahme eines Wörterbuches schreiben.
C2	Ich kann so gut schreiben, dass Muttersprachler meine Texte nicht überprüfen müssen.
C2	Ich kann so gut schreiben, dass meine Texte nicht wesentlich verbessert werden können, nicht einmal von einem Lehrer, der speziell das Schreiben übt.

Als Teil eines langfristigen Forschungsprojektes der Association of Language Testers in Europe (ALTE) wurden zusätzlich mit 400 Items weitere Kann-Beschreibungen in den drei Bereichen Soziales & Tourismus, Beruf und Bildung skaliert und mit dem Rasch-Modell den Niveaus des Referenzrahmens des GER angepasst (Europarat, 2001, S. 232-244).

Bei der Strukturierung der Sprachkompetenzen in Mündlichkeit (Hören/Sprechen) versus Schriftlichkeit (Lesen/Schreiben) muss bei einer differenzierteren Theoriebildung beachtet werden, dass sowohl die rezeptiven als auch die produktiven Kompetenzen jeweils nicht nur rein mündlich (Sprechen – Hören; Hören – Sprechen) respektive rein schriftlich (Lesen – Schreiben; Schreiben – Lesen) stattfinden, sondern auch in Kombination auftreten: sowohl mündlich – schriftlich (Hören – Schreiben; Sprechen – Schreiben) als auch schriftlich – mündlich (Lesen – Sprechen; Schreiben – Sprechen) (Hornung, 2005, S. 128). Die durch diese Kombination entstehende gegenseitige Beeinflussung – respektive Abhängigkeit[13] – wird beispielsweise bei der GER-Subskala „Notizen machen (in Vorträgen, Seminaren etc.)“ insofern berücksichtigt, als dass diese Skala nicht eindeutig der schriftlichen Produktion zugeordnet ist, sondern innerhalb der Betrachtung der wechselseitigen Beeinflussung von Text und Medium als Reaktion auf einen mündlichen Input und deren psycholinguistischen Konsequenzen geortet wird (Europarat, 2001, S. 95-99).

Für EVAMAR II können diese allenfalls weiter vorzunehmenden Differenzierungen vernachlässigt werden, weil sowohl die output-orientierte Leistungsmessung in Teilprojekt C als auch die Analysen der Maturitätsprüfungen in Teilprojekt D1 auf Prozessen reiner Schriftlichkeit beruhen (Lesen – Schreiben bzw. Textrezeption – Textproduktion).

2.2.4.2 Einschätzung des GER für die Verwendung in EVAMAR II

Eine Eingliederung in die europäische Sprachen-Gemeinschaft erscheint sinnvoll und wurde mit dem Schweizer Forschungsprojekt von Schneider & North (1999) zur Mitentwicklung am Gemeinsamen Europäischen Referenzrahmen (GER) bereits weit vorangetrieben. Der Europäische Referenzrahmen für Sprachen bietet ein flexibles Verzweigungssystem, durch welches Institutionen die für sie relevanten „Äste“ differenzieren und verfeinern und ihnen dabei doch einen Ort innerhalb des Gemeinsamen Referenzrahmens zuweisen können (Schneider & North, 1999, S. 16; Europarat, 2001, S. 40-42). So hat sich auch DESI als eine der grössten deutschlandweiten Schulleistungsstudien (siehe Abschnitt 2.2.2) mit dem Gemeinsamen Europäischen Referenzrahmen für Sprachen (GER) verlinkt und das Rahmenkonzept von GER ebenfalls bei der Entwicklung ihrer Testaufgaben berücksichtigt (Beck & Klieme, 2007, S. 2, S. 6; Nold, 2007; Takala, 2007).

Der Europäische Referenzrahmen deckt (im Unterschied zu PISA) den gesamten Bereich der Sprachkompetenzen ab: Lesen und Hören (als Verstehen), Sprechen und Schreiben – also sowohl Rezeption und Produktion als auch Interaktion – sowie als weitere kommunikative Sprachkompetenzen solche, welche der Sprachreflexion zugeordnet werden können. Sämtliche Bereiche wurden in mehreren, (meist) empirisch kalibrierten Subskalen ausdifferenziert. Die (produktive) Mündlichkeit wird dabei stark gewichtet, dies ist für das Sprachenlernen relevant, für das Studium jedoch nicht primär; dieser Einzel-Bereich kann für EVAMAR II zurückgestellt werden.

Für EVAMAR II konnte aufgrund unserer Einschätzung grundsätzlich an das Profil des Gemeinsamen Europäischen Referenzrahmens für den zu untersuchenden Bereich von Niveau C „Kompetente Sprachverwendung“ angeknüpft werden. Es mussten gewisse Adaptionen und die Ergänzung um die wissenschaftliche Textsorte „universitäre Fachtexte“ vorgenommen werden. Zudem hatte eine Klärung der zu untersuchenden Kompetenzbereiche zu erfolgen.

[13] Der vorgängige rezeptive Akt des Zuhörens muss erfolgreich gelingen, damit darauf basierend die zusammenfassende oder pointierte Verschriftlichung reüssieren kann; die dazu notwendigen Prozesse der Informations-Entnahme und -Gewichtung sind als weitere kognitive Teilkompetenzen zu erfassen.

Weiterhin galt es zu prüfen, ob eine auf GER aufbauende Zusatzstufe C3 oder D für universitäre bzw. wissenschaftliche Sprachverwendung aufgestockt und eingeführt werden könnte. Diese letztere Option wurde durch das Expertenteam der Projektpartner eindeutig verworfen.

Die Niveau-Kannbeschreibungen der Stufe C konnten für EVAMAR II als Grundlage übernommen werden, um darauf aufzubauen und sie mit Kompetenzbeschreibungen für die wissenschaftliche Textsorte zu ergänzen.

Da der Referenzrahmen einen „fundierten Lösungsversuch für dieses vorrangige und grundlegende Problem [der Vergleichbarkeit, die Autoren] des Lernens moderner Sprachen in einem europäischen Kontext" (Europarat, 2001, S. 177) bietet, erwies es sich als folgerichtig, sich mit den Untersuchungen für EVAMAR II innerhalb dieses erweiterbaren Systems zu verorten.

2.2.5 HarmoS

2.2.5.1 Kurzbeschreibung

Während der Laufzeit des Projekts EVAMAR II befand sich auch das Projekt der Entwicklung eines für die Volksschule der Schweiz verbindlichen Kompetenzmodells für Erstsprache im Rahmen des Projekts „Harmonisierung der obligatorischen Schule" (HarmoS) in einer fortgeschrittenen Phase. Weil es Grundlage verbindlicher Bildungsstandards für die Volksschule sein wird, ist eine Überprüfung der möglichen Verwendbarkeit im Projekt EVAMAR II sinnvoll.

Eines der Ziele der „Interkantonale[n] Vereinbarung über die Harmonisierung der obligatorischen Schule" (HarmoS-Konkordat) ist die Beschreibung von Kompetenzen und die Entwicklung von Bildungsstandards als Instrumente der Systementwicklung und Qualitätssicherung (EDK, 2007, Art. 7); dies erfolgt im Rahmen wissenschaftlicher Arbeiten innerhalb des Projektes HarmoS (EDK, 2004a, 2004b, 2007). In einem weiteren Schritt sollen die Lehrpläne, Lehrmittel, Evaluationsinstrumente und Bildungsstandards aufeinander abgestimmt werden (EDK, 2007, Art. 8).

Die im Rahmen von HarmoS zu entwickelnden und empirisch zu validierenden Kompetenzmodelle sollen differenzierte Beschreibungen von Kompetenzniveaus umfassen und es erlauben, auf politischer Ebene die Leistungsstandards pro Fachbereich per Ende des zweiten, sechsten und neunten Schuljahres festzulegen. Theoretisch basieren die Zielsetzungen und Konzeptionen von HarmoS auf der Kompetenzdefinition Weinerts (1999, 2001) sowie auf der Klieme-Expertise und deren Kompetenz-Modell-Bestimmung (Klieme et al., 2003, S. 74) (EDK, 2004a, S. 8, S. 9; Sieber, 2006b).

Der Leiter des Erstsprache-Konsortiums beschreibt die Schritte auf dem Weg zu einem HarmoS-Kompetenzmodell für Erstsprache wie folgt (Sieber, 2006a, 2006b): Die Komplexität sprachlichen Handelns wird matrixförmig übersichtlich reduziert, erstens durch eine Gliederung nach fünf Aspekten (Situieren, Planen, Realisieren, Evaluieren, Reparieren; Ebene 1) und zweitens mittels einer Unterscheidung der vier Modalitäten (Hören, Sprechen, Lesen, Schreiben; Ebene 2) (siehe Abbildung III.7). In den daraus entstehenden Feldern erfolgen konkrete Aufgabenbeschreibungen (Ebene 3) (Sieber, 2006a, S. 15; siehe Abbildung III.7).

Abb. III.7: HarmoS: Ein Modell für die Beschreibung (Sieber, 2006a, S. 18)

		Teilbereiche des sprachlichen Handelns (Ebene 2)			
		Hören	Sprechen	Lesen	Schreiben
Aspekte von Sprachhandlungen (Ebene 1)	Situieren	Aufgaben (Ebene 3)	Aufgaben	Aufgaben	Aufgaben
	Planen	Aufgaben	Aufgaben	Aufgaben	Aufgaben
	Realisieren	Aufgaben	Aufgaben	Aufgaben	Aufgaben
	Reparieren	Aufgaben	Aufgaben	Aufgaben	Aufgaben
	Evaluieren	Aufgaben	Aufgaben	Aufgaben	Aufgaben

Wie in HarmoS die vielfältigen Anforderungen und Teilkompetenzen in den jeweiligen Sprachbereichen weiter aufgegliedert werden, soll beispielhaft für den Sprachbereich Hören in Abbildung III.8 gezeigt werden.

Abb. III.8: HarmoS – ein Beispiel: Hören (Sieber, 2006a, S. 19)

Sprachhandlung	Hören
Situieren	Kommunikationsfunktion erkennen Kommunikationsform identifizieren Kommunikationsstile identifizieren (...) Kommunikationsziel im Auge behalten Gehörtes mit Vorwissen verbinden Motivation und Interesse aktivieren und aufrecht erhalten
Planen	Redeplan identifizieren Zuhörstrategien auswählen Sprachliches und nicht sprachliches Wissen strukturieren
Realisieren	Wissen aktivieren (Sprach- und Weltwissen) (...), sich auf die Äusserungen der Sprechenden konzentrieren Zuhörregeln beachten Verbale (...) Äusserungen/Geräusche verstehen; Kohärenz herstellen, Sinn konstruieren (...) Hilfsmittel benutzen (notes, tableau ...)
Reparieren	Bei Unverständnis wenn möglich Hören wiederholen (Tonträger) (...) Das Textverständnis rekapitulieren
Evaluieren	Permanentes Monitoring des Zuhörprozesses Wirkung der fremden Redeweise einschätzen Lokale und globale Sinnkonstruktion überprüfen, anpassen Verständnisfragen stellen, nachfragen, Begründungen verlangen Über Inhalt und Form reflektieren

Den unterschiedlichen Anforderungen, welche Texte an die Lesekompetenz stellen, wird HarmoS einerseits gerecht mit einer Berücksichtigung der Textformate „kontinuierlich versus diskontinuierlich“ (siehe auch Kapitel 2.2.3 zu PISA) und andererseits mit der Unterscheidung von Textarten (genres) und ihren unterschiedlichen Funktionen. Dabei wird im Projekt HarmoS differenziert zwischen (literarisch-)erzählenden, informatorischen und argumentativen Texten (Sieber, 2006c, S. 24-25).

2.2.5.2 Einschätzung des HarmoS-Modells für die Verwendung in EVAMAR II

Das Kompetenzmodell aus dem Projekt HarmoS wurde vor allem als Variante zu einer Verankerung des EVAMAR-Kompetenzrasters im Europäischen Sprachenrahmen GER (siehe Abschnitt 2.2.4) diskutiert. Aufgrund der folgenden Überlegungen konnte es nicht direkt übernommen werden: Gemäss den im ersten Abschnitt (III, 1) skizzierten, für EVAMAR II relevanten Aspekten wie hohe Jahrgangsstufe, Schultypus Gymnasium und hohe Kompetenzerwartung blieb es zweifelhaft, ob eine Adaption der EVAMAR-Testung an die drei Messzeitpunkte von HarmoS (Ende des zweiten, des sechsten und des neunten Schuljahres) möglich ist. Die wissenschaftliche Textsorte der universitären Fachtexte, auf der die EVAMAR-Leistungstests Erstsprache basieren sollten, war ein weiteres wichtiges Kriterium, welches einen Anschluss an das HarmoS-Projekt schwierig machte.

Wegen der grossen vorgymnasialen Bedeutung des HarmoS-Modells wurde auch eine mögliche Synthese zwischen dem „volksschulischen“ HarmoS-Modell und dem „tertiären“ GER geprüft. Dazu wurde aufgrund der Anregung der Projektpartner aus der Romandie und dem Tessin eine Variante eines Erstsprachen-Modells mit dem Ziel entwickelt, die Kompetenz-Teilbereiche nach GER matrixförmig mit den Sprachhandlungen nach HarmoS zu kombinieren. Es war der Versuch, die Kann-Beschreibungen beider Systeme zu analysieren und in ein sinnvolles neues Modell zu integrieren. Im Ergebnis erwiesen sich die dabei entstandenen Modell-Zellen vor allem deshalb als wenig bis kaum brauchbar, weil beide Kategorisierungen bereits in sich konsistent und somit nicht sinnvoll zu trennscharfen Zellen kombinierbar waren. Der Versuch wurde daher abgebrochen.

2.2.6 Lernstandserhebung in Nordrhein-Westfalen (NRW)

2.2.6.1 Kurzbeschreibung

Als weitere Orientierungsmöglichkeit bei der Entwicklung eines Komptenzrasters für Erstsprache im Projekt EVAMAR II bot sich das den Lernstandserhebungen in Nordrhein-Westfalen (NRW) zugrunde liegende Modell an. Seit dem Herbst 2004 sind dort für die jeweils rund 200’000 Schülerinnen und Schüler aller neunten Klassen – seit dem Schuljahr 2006/2007 aller achten Klassen – aller Haupt-, Real- und Gesamtschulen sowie Gymnasien die so genannten „landesweiten Lernstandserhebungen“ unter der Federführung des Ministeriums für Schule und Weiterbildung des Landes Nordrhein-Westfalen (MSWNRW) verpflichtend.

Leutner, Wirth und Fleischer (2005) betonen, dass es sich bei den Erhebungen nicht ausschliesslich um die Erfassung reiner Schulleistungen im Sinn von Performanz nach Prenzel, Drechsel, Carstensen und Ramm (2004) handelt, sondern die Kompetenzen in der Begriffsdefinition von Weinert (2001) als „Fähigkeitspotenziale“ – oder um mit Klieme et al. (2003, S. 72) zu sprechen als „befähigende Disposition“ – der Schülerinnen und Schüler zu sehen sind (S. 13).

Die fachspezifischen Kompetenzmodelle, aus denen die Aufgaben entwickelt wurden, erlauben die Zuordnung der Aufgaben zu bestimmten Kompetenzniveaus mit unterschiedlichen Schwierigkeitsgraden. Im Folgenden werden die Kompetenzniveaus für Deutsch angeführt, welche in den getesteten Bereichen der Lernstandserhebungen definiert wurden (MSWNRW, 2005b).

Im ersten Bereich – Leseverstehen (D-LV)/Reflexion über Sprache – wurden fünf Kompetenzniveaus gebildet: Das erste Niveau D-LV 1 wird als einfache Informationsentnahme (MSWNRW, 2005b) bzw. Leseverstehen in Ansätzen (MSWNRW, 2007) beschrieben, das zweite Niveau (D-LV 2) meint einfaches Leseverstehen, D-LV 3

bedeutet grundlegendes Leseverstehen, D-LV 4 differenziertes Leseverstehen und die höchste Stufe (D-LV 5) ist das komplexe Leseverstehen (MSWNRW, 2005b, S. 1-3; MSWNRW, 2007).

Der zweite Bereich – Zuhören und Verarbeiten (D-ZV) – wird in drei Stufen plus eine Zusatzstufe gegliedert: Kompetenzniveau D-ZV 1 entspricht der Leistung „einfache Einzelinformationen aus Hörbeiträgen erfassen und wiedergeben“, auf Niveau D-ZV 2 kann eine Schülerin oder ein Schüler „übersichtlich präsentierte Informationen aus Hörbeiträgen erfassen, identifizieren und zuordnen“, auf Kompetenzniveau D-ZV 3 „Informationen differenziert erfassen und komplexe Verarbeitungsleistungen erbringen“ und auf Kompetenzniveau D-ZV 3 plus kann eine Person, „Informationen selbstständig und differenziert erfassen und eigenständig verarbeiten“ (MSWNRW, 2005b, S. 4-5).

Im dritten Bereich – Schreiben (D-S) – meint das erste Kompetenzniveau (D-S 1) die elementare Erfassung und Darstellung des Sachverhaltes oder Themas, das zweite Niveau (D-S 2) die zielorientierte und aufgabenbezogene Darstellung des Sachverhaltes/Themas, und auf der höchsten Stufe (D-S 3) besitzen die Schülerinnen und Schüler die Fähigkeit der bewussten und zielgerichteten Gestaltung eines komplexen Schreibprozesses einschliesslich eigener Wertungen und Begründungen (MSWNRW, 2005b, S. 5-6).

Für EVAMAR II kam nur ein Bezug auf die obersten Kompetenzniveaus in Frage. Abbildung III.9 zeigt eine entsprechende Zusammenstellung. Die Matrix ist mit möglichen Kann-Beschreibungen gefüllt. Dem bei den Lernstandserhebungen NRW im Leseverstehen integrierten Teilbereich „Reflexion über Sprache“ wurde eine eigene Unterspalte zugewiesen.

2.2.6.2 Einschätzung der Lernstandserhebungen NRW für die Verwendung in EVAMAR II

Ausgehend von den unter Abschnitt 2.1 beschriebenen Aspekten bzw. Kriterien der hohen Jahrgangsstufe der Schülerinnen und Schüler bei EVAMAR II und der Fokussierung auf das Gymnasium – die Lernstandserhebungen NRW umfassen alle Schultypen –, wurde die Möglichkeit geprüft, sich allenfalls ausschliesslich auf das oberste Niveau der Kompetenz-Modellierungen der getesteten Deutschbereiche in den Lernstandserhebungen NRW zu fokussieren. Eine genauere Betrachtung ergab in der Folge, dass die vom Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (MSWNRW) gewählten drei Deutschbereiche „Leseverstehen/Reflexion über Sprache“, „Zuhören und Verarbeiten“ sowie „Schreiben“ sich mit ihren Restriktionen für die Testkonstruktion bei EVAMAR II nicht als sinnvoll erwiesen: Einerseits war im Projekt EVAMAR II zu beachten, dass sich die Bereiche Schreiben und (Zu-)Hören in der zur Verfügung stehenden Testzeit nicht umfassend prüfen liessen und die Korrekturen zu aufwändig gewesen wären. Andererseits stellte sich das Problem, dass die in NRW jährlich neu erstellten Aufgaben anhand interner Standards und die deshalb nicht gewährleistete längsschnittliche Vergleichbarkeit (MSWNRW, 2005b, S.1) das Herausarbeiten eines allgemein verwendbaren Modells erschwerten, wenn nicht gar verunmöglichten.

Abb. III.9: Die höchsten Kompetenzniveaus der Lernstandserhebungen Nordrhein-Westfalens 2005 – nachgebildete Matrix

Getestete Bereiche NRW 9	Leseverstehen/Reflexion über Sprache		Zuhören und Verarbeiten	Schreiben
Kompetenzniveau	D-LV 5: Komplexes Leseverstehen Schwierige Verarbeitungsprozesse werden bei der Lesekompetenz geleistet: Neben der Erfassung des ganzen Textes werden Inhalt und Struktur des Textes genauer betrachtet und beurteilt. Text-merkmale und ihre Auswirkungen werden differenziert erläutert.		D-ZV 3 plus: (Gehörte) Informationen selbstständig und differenziert erfassen und eigenständig verarbeiten	D-S 3: Bewusste und zielgerichtete Gestaltung eines komplexen Schreibprozesses einschliesslich eigener Wertungen und Begründungen
Beschriebene Teilkompetenzen	Finden und Verbinden auch von schwer zu ermittelnden Informationen	Kenntnisse und Anwendung grammatikalischer Grundbegriffe in vorgegebenen Kontexten	Bei der Verarbeitung eines (Radio-)Beitrags selbständige Schlussfolgerungen ziehen und klar darlegen können	Inhalte, Ideen und Aussagen – im Sinne der Aufgabenstellung – mehrdimensional anlegen und sachlich nachvollziehbar darstellen

Getestete Bereiche NRW 9	Leseverstehen/Reflexion über Sprache		Zuhören und Verarbeiten	Schreiben
	Genaues Verständnis der einzelnen Textabschnitte und des Textzusammenhanges	Bewertung und differenzierte Beurteilung von Inhalt und Struktur sprachlicher Phänomene	Auch nur kurz angesprochene Aspekte mit Blick auf die Aufgabe beachten können	Unterschiedliche Gesichtspunkte einbringen, begründen und abwägen
	Erkennen und Erklären von Mehrdeutigkeiten in Texten	Heranziehen und Anwenden von Fachwissen und Konzepten zum Verständnis und zur Beurteilung eines Textes	Den verarbeiteten Sachverhalt situationsbezogen darstellen können	Eigene Ideen, Gedanken den verschiedenen Aspekten der Aufgabenstellung anpassen
	Gesichertes Erarbeiten der Bedeutung des Gesamttextes	Reflexion über Inhalt, Sprache und Struktur des Gesamttextes		Schreibstrategien gezielt und der Aufgabenstellung angemessen einsetzen
	Klare Einordnung einer Aussage eines komplexen Textes in den Gesamtzusammenhang	Abwägen von Deutungsansätzen in Abgrenzung von anderen naheliegenden Möglichkeiten		
	Explizite Überprüfung von Deutungshypothesen anhand konkreter Textaussagen	Integration von Alltagswissen bzw. aussertextlichem Weltwissen bei der Erarbeitung des Textverständnisses und der kritischen Bewertung des Textes		
	Erfassen der Gesamtaussage eines komplexen literarischen oder Sachtextes			

2.3 Das Endmodell „Kompetenzraster Erstsprache" für EVAMAR II

Für die endgültige Auswahl von Komponenten aus den unter Abschnitt 2.2 beschriebenen Kompetenzmodellen und -rastern als Grundlage des Endmodells waren die in den Abschnitten 1 und 2.1 erläuterten Aspekte und Kriterien zu beachten. Die entsprechenden Einschätzungen finden sich in den jeweiligen Unterkapiteln der beschriebenen Studien. Zumindest einzelne Teile des GER entsprechen den Kriterien am besten (siehe die Einschätzung in Abschnitt 2.2.4.2). Im Folgenden wird die Begründung der Wahl des GER noch differenziert.

2.3.1 Begründung der Wahl des GER als Grundlage

Der mit dem Gemeinsamen Europäischen Referenzrahmen für Sprachen (GER) gewählte Anschluss-Referenzrahmen geniesst eine hohe Akzeptanz unter Hunderttausenden von Sprachlernenden und -Lehrenden im Bereich des lebenslangen Lernens sowie unter Wissenschaftlern und Didaktikerinnen in ganz Europa. Seine Subskalen sind empirisch kalibriert und werden in der Praxis laufend überprüft und verfeinert. Sowohl wegen seiner umfassenden Ausrichtung (Erfassung aller Teilbereiche der Sprachkompetenzen) als auch aufgrund seiner expliziten Ausbaufähigkeit hat sich der Gemeinsame Europäische Referenzrahmen für Sprachen unter allen untersuchten Sprachmodellen als herausragend geeignet für den Anschluss eines höheren Niveaus einer universitären Ausrichtung im Sinn von EVAMAR II erwiesen.

Da zudem die klassischen grossen Lern- und Aufgabenbereiche des Sprach- und somit Deutschunterrichtes je nach Didaktikauffassung und Zeitphase (Überblick beispielsweise Boueke[14], 1979) in drei (beispielsweise Ingendahl[15], 1999) respektive fünf Lernfelder (beispielsweise Glinz[16], Glinz & Ramseier, 1997; Abraham[17], Beisbart,

[14] Die wissenschaftliche Diskussion um den Deutschunterricht in den 60er-Jahren und dessen Veränderungen zeichnet beispielsweise Boueke (1979) mit Forschungsberichten von 1965 bis 1977 nach.

[15] Ingendahl (1999) unterscheidet als „in sprachdidaktischer Theoriebildung, aber auch in Lehrplänen und Schulbüchern weitgehend üblich" die drei muttersprachlichen Aufgabenbereiche 1. sprechen und schreiben zu lehren, 2. Äusserungen und Texte verstehen zu lehren, 3. über Sprachliches reflektieren zu lehren. Der dritte Bereich, der früher schlicht Grammatikunterricht hiess, wird „offiziell ... schon seit einigen Jahrzehnten" Sprachreflexion oder „Reflexion über Sprache und Sprachgebrauch" ge-

Koß & Marenbach, 2003, S. 9) eingeteilt werden und bezüglich dieser unterschiedlichen Differenzierung in der Fachdidaktik Deutsch weiterhin kein Konsens besteht (Ossner, 2006, S. 7)[18], erscheint für unseren Bedarf eine Bezugnahme auf die im europäischen Sprachraum akzeptierte Lernfeldbetrachtung des GER als einigendes Instrument adäquat.

Unsere Bemühungen gingen infolgedessen dahin, den Referenzrahmen für Sprachen von GER für die Bedürfnisse des Projektes EVAMAR II zu adaptieren, indem er beispielsweise mit der Textsorte „wissenschaftliche Fachtexte" um die universitäre Stufe erweitert wurde. Die konkreten Beschreibungen dazu finden sich in Kapitel 2.3.2.

Eine Kombination mit den HarmoS- und PISA-Kompetenzmodellen wurde zwar diskutiert, dann aber aus verschiedenen Gründen wieder verworfen (siehe Abschnitt 2.2.5.2). Die Variante, die Subskalen des Teilbereiches Lesen nach PISA 2000 aufgrund der drei empirisch elaborierten Berichtskalen „Informationen ermitteln", „Textinterpretation" und „Reflektieren & Bewerten" einzubauen, wurde auf Beschluss des Projektteams nicht weiter verfolgt.

2.3.2 Beschreibung des EVAMAR-Kompetenzrasters für Erstsprache

In das für EVAMAR II entworfene, übergeordnete Modell für Erstsprache wurden aus dem GER die wichtigsten unterscheidenden Teil-Skalen übernommen, welche in den ersten Universitätssemestern relevant sind und unterschiedliche Fähigkeiten und Fertigkeiten erfordern. Verzichtet wurde beispielsweise auf Beschreibskalen wie „kreatives Schreiben", „Korrespondenz lesen" u. ä.

Auf der horizontalen Achse übernimmt die Sprachen-Matrix grundsätzlich die Europäische Sprachen-Niveau-Stufung des GER, jedoch nur den für Maturandinnen und Maturanden sowie Studierende relevanten Bereich der kompetenten Sprachverwendung C sowie die obere Hälfte der selbstständigen Sprachverwendung B. Die unterste Stufe B 2.1 wurde mitberücksichtigt, weil sie an Deutschschweizer Universitäten als Minimalkompetenz in Deutsch als Fremdsprache (DaF) gilt, welche ausländische Studierende mit anderer Muttersprache zum Nachweis ihrer Studierfähigkeit in der Studiensprache Deutsch für die Zulassung an die Schweizer Universitäten in der entsprechenden Prüfung zeigen müssen. Die Stufe C 1.1 gilt als zu erreichendes Test-Maximum der Universitäts-Zulassungsprüfung für ausländische Studierende mit Deutsch als Fremdsprache. Der erwartete Anforderungsbereich für Studierende mit Deutsch als Muttersprache liegt auf Niveau C. Die Überlegung, eine Zusatzstufe C3 oder D für universitäre bzw. wissenschaftliche Sprachverwendung einzuführen, wurde durch das Expertenteam der Projektpartner eindeutig verworfen. Somit befinden sich auf der Horizontalachse sechs Niveaustufen von B 2.1 bis C 2.2.

Als Grundmodell für Erstsprache ergibt sich damit eine zweidimensionale Matrix. Die Vertikalachse enthält die Sprachverwendungs-Lernbereiche Lesen, Hören, Schreiben, Reden und Sprachreflexion (mit Wortschatz, Grammatik und Orthografie), wobei jeder Bereich mit ausgewählten Teilkompetenzen weiter differenziert werden kann. Auf der Horizontalachse befinden sich auf gymnasialer und universitärer Ebene die drei bzw. sechs Niveaustufen B 2.1 bis C 2.2, sodass sich insgesamt mindestens 84 inhaltlich zu beschreibende Zellen ergeben (siehe Abbildung III.10).

---

nannt (S. 1). Ingendahl vernachlässigt hier die zunehmende Betonung der Differenzierung zwischen Mündlichkeit und Schriftlichkeit, womit aus seinen erwähnten drei Bereichen auch fünf würden.

16 Glinz, Glinz & Ramseier (1997) nehmen die fünf Bereiche „Sehen, Hören, Sprechen, Lesen, Schreiben " auf den Umschlag ihres Grundlagenbandes zum Lehrmittel „Das Schweizer Sprachbuch", orientieren sich im Aufbau jedoch wie folgt an den fünf Aufgabenfeldern des Sprachunterrichts: Sprechen und Handeln, Lesen und Verstehen, Texte schreiben, Rechtschreibung, Grammatik – wobei das Hören in das Feld „Sprechen und Handeln" integriert und die „Sprachbetrachtung" aufgeteilt wird.

17 Abraham et al. (2003) beschreiben den gegenwärtigen Stand der Diskussion für die traditionellen fünf Handlungs- oder Aufgabenfelder des Deutschunterrichts (Mündlicher Sprachgebrauch, schriftlicher Sprachgebrauch, Sprachbetrachtung, Rechtschreiben), sowie 62 ausgewählte *Tätigkeiten* (von Autorenlesungen bis *Z*usammenfassen) innerhalb dieser Lernbereiche.

18 Auch das jüngste, von Ossner (2006) mit dem Ziel der Konsensfindung entwickelte Kompetenzmodell bewirkte zwar eine Vertiefung der Diskussion, nicht aber eine breite Akzeptanz. Eine differenziertere Betrachtung der unterschiedlichen Schulen und Didaktiken oder der gegenwärtigen Handhabungen in den deutschen Bundesländern im Vergleich zur Schweiz würde hier den Rahmen sprengen.

Abb. III.10: Grundmodell Erstsprache nach GER (die von EVAMAR II untersuchten Bereiche sind schraffiert)

				Selbstständige Sprachverwendung		Kompetente Sprachverwendung			
				B 2		C 1		C 2	
				B 2.1	B 2.2	C 1.1	C 1.2	C 2.1	C 2.2
				Uni-Zulassung für DaF		Testmax. Unizul. DaF			
				EVAMAR II					
VERSTEHEN	Hören								
	Lesen	Leseverstehen allgemein	EVAMAR II						
		Zur Orientierung lesen							
		Information & Argumentation verstehen							
		Schriftliche Anweisungen verstehen							
SPRACHREFLEXION	Grammatische Kompetenz								
	Wortschatz	Wortschatz-Spektrum							
		Wortschatz-Beherrschung							
	Orthografie								
PRODUKTION	Schreiben	Berichte und Aufsätze schreiben							
		Schriftliche Produktion allgemein							
		Schriftliche Interaktion allgemein							
		Texte verarbeiten							
	Sprechen	Zusammenhängend sprechen							
		An Gesprächen teilnehmen							

Da von vornherein feststand, dass aufgrund der vorhandenen Ressourcen des Projektes und damit zusammenhängenden messtechnischen Problemen nicht sämtliche Teil-Kompetenzen empirisch untersucht werden konnten, welche für die Bewältigung der ersten Semester an einer Universität wesentlich sind, richtete sich das Hauptaugenmerk der Untersuchung im Feld der so genannten Sprachkompetenzen auf Teil-Kompetenzen, welche das Lesen betreffen. Diese grundlegende rezeptive Tätigkeit umfasst in der universitären Bildung die meiste Zeit des Lernens und bildet die Basiskompetenz, mithilfe derer neues Wissen angeeignet wird. Sind die Studierenden nicht in der Lage, die ausgeteilten oder elektronisch zur Verfügung gestellten Handouts, Skripts und Vorlesungsfolien im Detail richtig zu lesen und zu verstehen, die oftmals komplexen wissenschaftlichen Abhandlungen selektiv zu erfassen sowie Forschungsbeiträge in ihren Argumentationssträngen zu begreifen, so kann kein Lernen erfolgen, es findet keine Aneignung neuen Wissens statt.

Ebenso wichtig ist die Kompetenz des Hörens, welche in den Vorlesungen und Seminaren selbstverständlich auch eine nicht zu unterschätzende Rolle für das universitäre Lernen spielt. Sie gehört wie das Lesen zur rezeptiven Seite der Sprachmodelle, verläuft aber über einen anderen Kanal und wird zudem häufig von visuellen Medien/Reizen begleitet. Eine Untersuchung der Kompetenz des Hörens musste aus der Konzeption von EVAMAR II herausfallen, da sich das Untersuchungsdesign (Produktion von gestellten oder echten Video-Aufnahmen von Universitätsvorlesungen als Grundlage standardisierter Testitems) als sehr aufwändig erwies. Nichtsdestotrotz wäre es wünschenswert, auch diesen Zweig zu einem späteren Zeitpunkt auf universitärer Ebene zu testen.

Auf die Testung des Schreibens im Sinne des Verfassens wissenschaftlicher Arbeiten wurde für Teilprojekt C nicht nur aufgrund aufwändiger Korrekturverfahren verzichtet, sondern auch aus inhaltlichen Gründen: (Pro-)Seminararbeiten, also Textproduktionen mittlerer Länge, werden meist erst im späteren Verlauf des Studiums verlangt, sodass diese Fertigkeit nicht als Gegenstand im Untersuchungszeitraum von EVAMAR II erscheint.

Weiterhin konnten folgende sprachliche Teil-Kompetenzen nicht in das Untersuchungsfeld einbezogen werden: mündliche Kommunikation, Dialog- und Gesprächsverhalten sowie Vortragskompetenz und Präsentation, also monologisches und dialogisches Reden. Den im GER stark betonten Sprachverwendungsbereich der Interaktion konnte das universitär ausgerichtete Modell für EVAMAR vernachlässigen, ebenso zum Beispiel bei der Kompetenz „Sprechen“ die unterschiedlichen Teil-Kompetenzen „an Gesprächen teilnehmen“ oder „zusammenhängend sprechen“.

Da sich aus Teilprojekt A4 ergeben hatte, dass mehrere Dozierende die offensichtlich zunehmende Rechtschreibeschwäche der Studierenden bemängeln (siehe Abschnitt 3.3 in Kapitel II), wurden hingegen die sich in Textproduktionen zeigenden Orthografie- und Grammatik-Kompetenzen als eigenständige Bereiche getestet.

Somit beschränkt sich der Kompetenzraster Erstsprache in EVAMAR II vor allem im Hinblick auf die Tests in Teilprojekt C auf die Kompetenzbereiche Lesen und Sprachreflexion. Dies entspricht auch der Orientierung des Teilprojekts A1 am Textverständnis, also der Lese-Analyse vielfältigster Studienmaterialien.

Beim Bereich Lesen wurden die schwierig zu unterscheidenden Teilkompetenzen „Leseverstehen allgemein“ und „Zur Orientierung lesen“ aus testkonstruktionsbedingten Gründen zusammengefasst zum Kompetenzbereich „Allgemeines Leseverstehen“. In Gegenüberstellung dazu wurde die sich davon unterscheidende Lesefähigkeit „Information und Argumentation verstehen“ als „Detailliertes Leseverstehen“ bestimmt.[19] Für den Bereich Sprachreflexion wurden die zwei Teilkompetenzen „Wortschatz-Spektrum“ und „Wortschatz-Beherrschung“ zu einer umfassenderen Kompetenz „Wortschatz“ zusammengefasst, da sich bereits bei der Testkonstruktion Schwierigkeiten bei der Abgrenzung der beiden Subkategorien zeigten. Dieser zweite Schwerpunkt, der somit auf die Verwendung und die Geläufigkeit des (wissenschaftlichen) Wortschatzes gelegt wurde, entspricht eben-

[19] Auf den Teilbereich „schriftliche Anweisungen verstehen“ wurde im Hinblick auf die Tests verzichtet. Diese Teilkompetenz kann mit den bekannten PISA-Aufgaben wie „Gebrauchsanleitungen verstehen“ verglichen werden, ist in universitärem Rahmen der Erstsemestrigen aber eher selten als Anforderung gegeben.

falls der Fragestellung in Projekt A1, in welcher eine Erfassung und Codierung von Sinneinheiten des Wortschatzes als vorausgesetztes Wissen und Können erfolgte. Der dritte Schwerpunkt, der nur in der Erstsprache Deutsch getestet wurde (zur Begründung siehe Berichtsteil IV), umfasst die beiden Kompetenzen Grammatikkenntnisse und Orthografie. Zur Prüfung dieses letzten Bereiches wurde im Teil 10 des Deutschtests ein kürzeres offenes Format gewählt, in dem auch Elemente der freien Textproduktion zum Tragen kommen. Hier überlagern sich also die zwei Kompetenzbereiche Sprachreflexion und Produktion.

Dem resultierenden Kompetenzraster Erstsprache (siehe Abbildung III.11) wurde die spezifische Testsorten-Orientierung, die sich aus der Konzeption von EVAMAR II ergab, als horizontale Achse beigegeben: die wissenschaftlich universitäre Textsorte. Somit ergeben sich für das Submodell Erstsprache, das als Referenzraster und theoretische Verortung für die Testphase C dient, noch fünf inhaltlich zu beschreibende Zellen. Die Zelleninhalte bestehen aus Formulierungen von „Kann-Beschreibungen“: Es handelt sich dabei um Kürzungen und Verdichtungen aus den Beispielskalen des GER zum Niveau C2; ausgeklammert wurden für die universitäre Sprachverwendung und somit für EVAMAR II nicht relevante Bereiche wie z. B. umgangssprachliche und literarische Texte.[20] Für die wissenschaftliche Textsorte (universitäre Fachtexte) wurden für EVAMAR II neu konstruierte Formulierungen eingefügt (kursiv hervorgehoben).

2.4 Zusammenfassung

Die Literaturanalyse bestehender Sprachleistungsstudien und ihre theoretische Verortung in Sprachmodellen ergab ein heterogenes Bild. So verwendet beispielsweise die bisher grösste deutschdidaktische Studie „Deutsch-Englisch-Schülerleistungen-International“ (DESI) sechs Kompetenzbereiche mit drei bis fünf Kompetenzniveaus für neunte Klassen. Im Unterschied dazu bildet das Erstsprachenmodell in PISA für die 15-/16-Jährigen ausschliesslich den Teilbereich Lesen ab. Als eine der wenigen empirischen Studien an der Schnittstelle Gymnasium – Universität beschreibt das Zürcher Textanalyseraster hingegen ausschliesslich die schriftliche Produktion (Maturaaufsätze). Aufgrund der vergleichenden Analyse bestehender internationaler und nationaler Sprachmodelle wurde für EVAMAR II schliesslich ein Kompetenzraster für Erstsprache gewählt, der sich grundsätzlich am Gemeinsamen Europäischen Referenzrahmen für Sprachen (GER) (Europarat, 2001) orientiert. Damit knüpft das schweizerisch-nationale Projekt EVAMAR II am selben Referenzrahmen an wie die deutsch-nationale Studie DESI. Eine Verwendung von Komponenten aus den HarmoS- und PISA-Kompetenzmodellen wurde zwar diskutiert, aus verschiedenen Gründen aber schliesslich verworfen.

[20] Wie bereits erwähnt, war selbst in der Germanistik der Anteil der literarischen Texte, mit welchen Studierende in den ersten Semestern konfrontiert werden, anteilsmässig überraschend gering im Vergleich zu den theoretischen Einführungen und konnte somit vernachlässigt werden (siehe Projekt A1).

Abb. III.11: Kompetenzraster für die Tests in Erstsprache (mit Kann-Beschreibungen)

			Textsorten-Orientierung
			Universitäre Textsorte
VERSTEHEN	Lesen	Allgemeines Leseverstehen (Zur Orientierung lesen)	Kann alle Arten geschriebener Texte verstehen und kritisch interpretieren (einschliesslich abstrakter, strukturell komplexer, nicht-literarischer Texte). Kann ein breites Spektrum langer und komplexer Texte, *auch wissenschaftliche* Texte, verstehen und dabei feine stilistische Unterschiede und implizite Bedeutungen erfassen. Kann lange und komplexe Texte, *auch wissenschaftliche Texte,* rasch durchsuchen. Kann rasch den Inhalt und die Bedeutung von Artikeln und Texten zu einem breiten Spektrum *wissenschaftlicher* Themen erfassen und entscheiden, ob sich ein genaueres Lesen lohnt.
		Detailliertes Leseverstehen: Information & Argumentation verstehen	Kann ein weites Spektrum langer, komplexer Texte, *denen Studierende im ersten Studienjahr an der Universität begegnen*, verstehen und dabei feinere Nuancen auch von explizit oder implizit angesprochenen Einstellungen und Meinungen erfassen. Kann wichtige Einzelinformationen auffinden.[21]
SPRACHREFLEXION	Wortschatz		Beherrscht einen sehr reichen Wortschatz und ist sich der jeweiligen Konnotation bewusst. Durchgängig korrekte und angemessene Verwendung des Wortschatzes.
	Grammatische Kompetenz		Zeigt auch bei der Verwendung komplexer Sprachmittel eine durchgehende Beherrschung der Grammatik.
	Orthografie		Die schriftlichen Texte sind frei von orthografischen Fehlern.

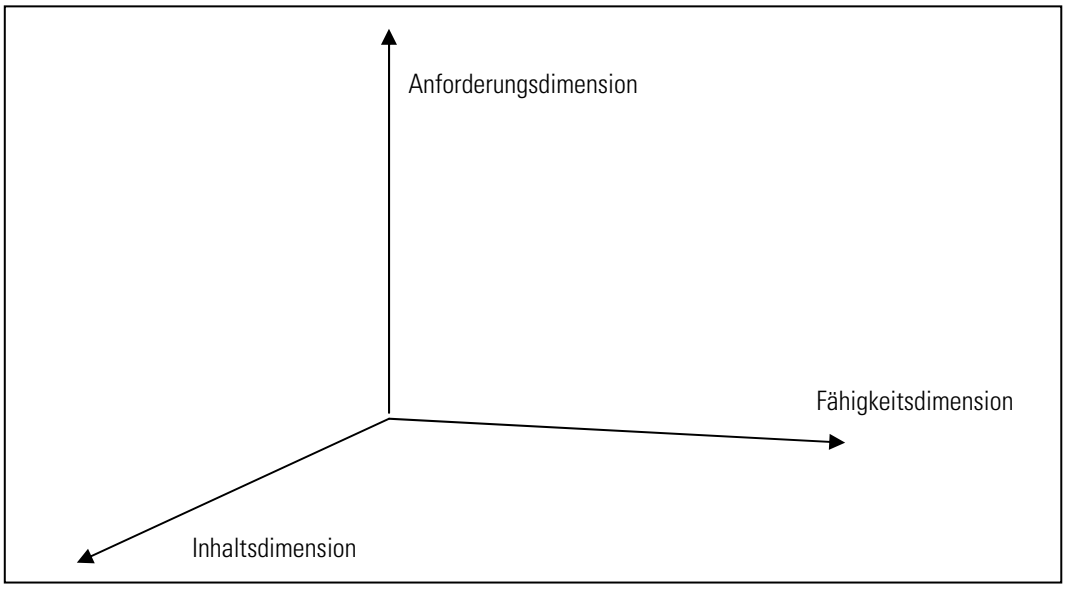
[21] Die Teilkompetenz „wichtige Einzelinformationen auffinden“ wurde für EVAMAR II aus dem Bereich des Allgemeinen zum Detaillierten Leseverstehen umgeordnet.

3 Der Kompetenzraster in Mathematik

3.1 Grundlagen

Für die theoretische Verortung des zu entwickelnden Kompetenzrasters für Mathematik wurden ebenfalls bisherige Arbeiten aus dem internationalen, europäischen, nationalen und auch regionalen Raum näher betrachtet. In Frage kamen Studien wie PISA, TIMSS oder HarmoS, in deren Rahmen Mathematik eines der untersuchten Schulfächer darstellt. Die Konstruktion der entsprechenden Mathematik(test)aufgaben stützt sich meist auf die Grundlage eines Kompetenzmodells. Kompetenzmodelle, die für mathematische Zwecke entwickelt wurden, sind zwei- oder mehrdimensionale Schemata, in denen Inhalte, Fähigkeiten und eventuell Anforderungsbereiche oder Kompetenzstufen festgelegt und beschrieben werden (siehe Abbildung III.12).

Abb. III.12: Mögliche Dimensionen eines Kompetenzmodells



Die Inhalts- und Fähigkeitsdimensionen werden häufig in einer Matrixform beschrieben (siehe Abbildung III.13).

Abb. III.13: Darstellungsform Matrix für die Inhalts- und Fähigkeitsdimensionen

		Inhaltsdimension				
		Inhalt 1	Inhalt 2	Inhalt 3		Inhalt n
Fähigkeits dimension	Fähigkeit 1					
	Fähigkeit 2					
	Fähigkeit 3					
	...					
	Fähigkeit m					

Im Folgenden sollen einige bekannte Kompetenzmodelle für Mathematik dargestellt und im Hinblick auf ihre Tauglichkeit für EVAMAR II analysiert werden. Sie unterscheiden sich vor allem in der Art, wie umfassend die Inhalte und/oder die Fähigkeiten und wie ausführlich die einzelnen Zellen der Matrix beschrieben werden.

3.2 Bestehende Kompetenzmodelle und Aufgabenraster für Mathematik

3.2.1 Das Kompetenzmodell des NCTM (National Council of Teachers of Mathematics)

Im Kompetenzmodell des NCTM werden fünf inhaltliche Dimensionen (Number & Operations, Algebra, Geometry, Measurement und Data Analysis & Probability) sowie fünf Fähigkeitsaspekte (Problem Solving, Reasoning & Proof, Communication, Connections und Representation) unterschieden (NCTM, 2000). Dieses in den USA entwickelte Modell diente einigen weiteren Kompetenzmodellen als Vorbild.

3.2.2 Das Kompetenzmodell der KMK in Deutschland

In Deutschland wurden Kompetenzmodelle unter anderem im Rahmen der Festlegung von Bildungsstandards zunächst für die 9. und 10. Jahrgangsstufe entwickelt (siehe Klieme et al., 2000).

Im Matrixmodell der deutschen Kultusministerkonferenz sind die beiden Dimensionen Inhalte (Leitideen) und allgemeine mathematische Kompetenzen zusammengefasst (siehe Abbildung III.14).

Abb. III.14: Das Kompetenzmodell der deutschen Kultusministerkonferenz

		Leitideen				
		Zahl	Messen	Raum und Form	Funktionaler Zusammenhang	Daten und Zufall
Allgemeine mathematische Kompetenzen	Mathematisch argumentieren					
	Probleme mathematisch lösen					
	Mathematisch modellieren					
	Mathematische Darstellungen verwenden					
	Mit symbolischen, formalen und technischen Elementen der Mathematik umgehen					
	Kommunizieren					

Eine Leitidee vereinigt Inhalte verschiedener mathematischer Sachgebiete und durchzieht ein mathematisches Curriculum spiralförmig. Die allgemeinen mathematischen Kompetenzen werden sodann inhaltlich nach Leitideen geordnet ausführlich beschrieben. Im Folgenden findet sich exemplarisch die Konkretisierung der ersten Leitidee (Zahl), eine genaue Zuordnung zu den allgemeinen mathematischen Kompetenzen fehlt allerdings:

Die Schülerinnen und Schüler

- nutzen sinntragende Vorstellungen von rationalen Zahlen, insbesondere von natürlichen, ganzen und gebrochenen Zahlen entsprechend der Verwendungsnotwendigkeit,
- stellen Zahlen der Situation angemessen dar, unter anderem in Zehnerpotenzschreibweise,
- begründen die Notwendigkeit von Zahlbereichserweiterungen an Beispielen,
- nutzen Rechengesetze, auch zum vorteilhaften Rechnen,

- nutzen zur Kontrolle Überschlagsrechnungen und andere Verfahren,
- runden Rechenergebnisse entsprechend dem Sachverhalt sinnvoll,
- verwenden Prozent- und Zinsrechnung sachgerecht,
- erläutern an Beispielen den Zusammenhang zwischen Rechenoperationen und deren Umkehrungen und nutzen diese Zusammenhänge,
- wählen, beschreiben und bewerten Vorgehensweisen und Verfahren, denen Algorithmen bzw. Kalküle zu Grunde liegen,
- führen in konkreten Situationen kombinatorische Überlegungen durch, um die Anzahl der jeweiligen Möglichkeiten zu bestimmen,
- prüfen und interpretieren Ergebnisse in Sachsituationen unter Einbeziehung einer kritischen Einschätzung des gewählten Modells und seiner Bearbeitung.

Zum Lösen mathematischer Aufgaben sind die allgemeinen mathematischen Kompetenzen in unterschiedlicher Ausprägung erforderlich. Es werden drei Anforderungsbereiche unterschieden:

- Anforderungsbereich I: Reproduzieren: Dieser Anforderungsbereich umfasst die Wiedergabe und direkte Anwendung von grundlegenden Begriffen, Sätzen und Verfahren in einem abgegrenzten Gebiet und einem wiederholenden Zusammenhang.
- Anforderungsbereich II: Zusammenhänge herstellen: Dieser Anforderungsbereich umfasst das Bearbeiten bekannter Sachverhalte, indem Kenntnisse, Fertigkeiten und Fähigkeiten verknüpft werden, die in der Auseinandersetzung mit Mathematik auf verschiedenen Gebieten erworben wurden.
- Anforderungsbereich III: Verallgemeinern und Reflektieren: Dieser Anforderungsbereich umfasst das Bearbeiten komplexer Gegebenheiten u. a. mit dem Ziel, zu eigenen Problemformulierungen, Lösungen, Begründungen, Folgerungen, Interpretationen oder Wertungen zu gelangen.

Im Allgemeinen nehmen Anspruchsniveau und kognitive Komplexität der Anforderungsbereiche in der angegebenen Reihenfolge zu.

3.2.3 Das Kompetenzmodell für Mathematik bei HarmoS

Die folgenden Ausführungen sind dem Zwischenbericht zu HarmoS (Linneweber & Wälti, 2006) entnommen.

Bei HarmoS wird ein Matrixmodell verwendet, das zwischen inhalts- und prozessbezogenen Komponenten unterscheidet. Im Gegensatz zu anderen Modellen werden die beiden Komponenten auch explizit miteinander verknüpft. Das, was Schülerinnen und Schüler können sollen, ist immer durch inhaltliche und prozess-/aktionsbezogene Komponenten bestimmt.

Das Kompetenzmodell von HarmoS umfasst vier Dimensionen: Entwicklungsdimension, Kompetenzaspekte, Kompetenzbereiche und Kompetenzniveaus (Schwierigkeitsniveaus).

In der Entwicklungsdimension werden Kompetenzen für die drei Jahrgangsstufen Ende

2. Schuljahr, Ende 6. Schuljahr und Ende 9. Schuljahr beschrieben.

Die acht Kompetenzaspekte beschreiben Fähigkeiten, die zum Lösen von Mathematikaufgaben nötig sind:

- Wissen, Erkennen und Beschreiben,
- Operieren und Berechnen,
- Instrumente und Werkzeuge verwenden,
- Darstellen und Formulieren,

- Mathematisieren und Modellieren,
- Argumentieren und Begründen,
- Interpretieren und Reflektieren der Resultate,
- Erforschen und Explorieren.

Die fünf Kompetenzbereiche beschreiben die folgenden inhaltlichen Gebiete der (Schul-) Mathematik:

- Form und Raum,
- Zahl und Variable,
- Funktionale Zusammenhänge,
- Grössen und Masse,
- Daten und Zufall.
- Das erste Gebiet umfasst geometrische, das zweite algebraische, das dritte funktionale Sachverhalte, das vierte Gebiet lässt sich auch mit Sachrechnen und das fünfte mit Stochastik umschreiben.

Die Kompetenzniveaus (I, II, III, IV) werden durch Aufgaben unterschiedlicher Schwierigkeit konkretisiert und erst empirisch über die Ergebnisse von Tests endgültig festgelegt. Der Schwierigkeitsgrad von Aufgaben wird bezüglich verschiedener Kategorien beschrieben:

- Schwierigkeit bezüglich der Aufgabenstellung
- Schwierigkeit bezüglich der Aufgabenbearbeitung (Komplexität)
- Schwierigkeit bezüglich der Bekanntheit der Aufgabe (Lehrplan)

Die acht Kompetenzaspekte angewendet auf die fünf Kompetenzbereiche führen zu einer Matrix mit 40 (zunächst einmal leeren) Feldern (siehe Abbildung III.15 für die Jahrgangsstufe 9). Für jedes der 40 Felder wird beschrieben, was Schülerinnen und Schüler der betreffenden Jahrgangsstufe können sollen. Diese „Can do's" bilden zusammen mit konkreteren Auflistungen von Fähigkeiten und Fertigkeiten die Basis für Aufgaben resp. Items auf vier verschiedenen Schwierigkeitsniveaus.

Abb. III.15: Kompetenzaspekte und -bereiche bei HarmoS (9. Jahrgangsstufe)

		Inhaltsdimension („Kompetenzbereiche")				
		Zahl und Variable	Form und Raum	Grössen und Masse	Funktionale Zusammen-hänge	Daten und Zufall
Prozessdimension („Kompetenzaspekte")	Wissen, Erkennen und Beschreiben					
	Operieren und Berechnen					
	Instrumente und Werkzeuge verwenden					
	Darstellen und Formulieren					
	Mathematisieren und Modellieren					
	Argumentieren und Begründen					
	Interpretieren und Reflektieren der Resultate					
	Erforschen und Explorieren					

3.2.4 Das Kompetenzmodell für Mathematik bei PISA

Das Kompetenzmodell bei PISA ist dreidimensional (Klieme, Neubrand & Lüdtke, 2001/2002). Unterschieden werden die folgenden Bereiche: Inhalte, hierarchisch gegliederte Fähigkeiten und Kompetenzstufen.

Die Inhaltsbereiche sind:

- Quantität
- Veränderung und Beziehung
- Raum und Form
- Unsicherheit

Die mathematischen Fähigkeiten sind:

- Kommunizieren
- Modellieren
- Argumentieren
- Technisch arbeiten
- Repräsentieren
- Problemlösen

Diese Bereiche und Fähigkeiten lehnen sich weitgehend an das Modell des NCTM an (siehe Abschnitt 3.2.1).

In jedem Leistungsbereich, das heisst in jeder Zelle der Inhalts-/Fähigkeitsmatrix werden fünf Kompetenzstufen bzw. -niveaus unterschieden:

- Stufe I: Rechnen auf Grundschulniveau
- Stufe II: Elementare Modellierungen
- Stufe III: Modellieren und begriffliches Verknüpfen auf dem Niveau der Sekundarstufe I
- Stufe IV: Umfangreiche Modellierungen auf der Basis anspruchsvoller Begriffe
- Stufe V: Komplexe Modellierung und innermathematisches Argumentieren

Die Beschreibungen der Kompetenzstufen enthalten Aussagen auf vier inhaltlichen Ebenen (siehe Klieme, Neubrand & Lüdtke, 2001, S. 160):

- Wissen und Wissensinhalte
- Modellierungsanforderungen
- Verknüpfung von Konzepten
- Offenheit der Aufgabe

3.2.5 Modelle bei TIMSS

Im Rahmen der verschiedenen TIMSS-Studien wurden unterschiedliche Kompetenzmodelle entwickelt. Die hier besonders relevante Untersuchung ist TIMSS/III, bei der die mathematischen und physikalischen Kompetenzen am Ende der gymnasialen Oberstufe untersucht wurden und die sich an „gymnasialer Mathematik und Physik“ (Ramseier et al., 1999, S. 130) orientiert. Ein Aspekt ist der Versuch, schwierigkeitsgenerierende Merkmale von Mathematik- und Physikaufgaben zu bestimmen.

Für die TIMSS/III-Studie wurde ein im Vergleich zu unteren Klassenstufen recht differenziertes Modell verwendet. Eine ausführliche Beschreibung des Modells findet sich bei Baumert, Bos & Lehmann (2000).

Es enthält vier Stufen von Verhaltenserwartungen:

- Wissen
- Beherrschen von Routineverfahren
- Beherrschen von komplexen Verfahren
- Anwendungsbezogene Aufgaben und innermathematische Probleme

Bezüglich der Anspruchsniveaus werden ebenfalls vier Stufen unterschieden:

- Stufe 1: Alltagsbezogene Schlussfolgerungen
- Stufe 2: Anwendung von einfachen Routinen
- Stufe 3: Bildung von Modellen und Verknüpfung von Operationen
- Stufe 4: Mathematisches Argumentieren

Die mathematischen Inhalte sind in fünf Gebiete unterteilt:

- Zahlen, Gleichungen und Funktionen
- Analysis
- Geometrie, analytische Geometrie
- Wahrscheinlichkeit, Statistik
- Aussagenlogik, Beweise

Die Inhalte und die Verhaltenserwartungen lassen sich übersichtlich als Matrix darstellen (siehe Abbildung III.16).

Abb. III.16: Die Matrix der Verhaltenserwartungen und der Inhalte

		Verhaltenserwartungen			
		Wissen	Beherrschen von Routineverfahren	Beherrschen von komplexen Verfahren	Anwendungsbezogene Aufgaben und innermathematische Probleme
Inhalte	Zahlen, Gleichungen und Funktionen				
	Analysis				
	Geometrie, analytische Geometrie				
	Wahrscheinlichkeit, Statistik				
	Aussagenlogik, Beweise				

Für mathematische und naturwissenschaftliche Testaufgaben wurde bei TIMSS/III ein zusätzliches Kategoriensystem entwickelt. Darin sind acht gemeinsame Merkmale (Variablen) aufgeführt. Bei diesen Merkmalen handelt es sich um spezifische Anforderungen an Wissen, Verständnis, Denkprozesse und kognitive Grundfähigkeiten:

- G1. Kenntnis von Definitionen, mathematischen Sätzen (Müssen solche Wissensinhalte aus dem Gedächtnis abgerufen werden?)

- G2. Qualitatives Verständnis mathematischer Begriffe (Muss man beispielsweise verstanden haben, wie ein Fachbegriff mit anderen Begriffen zusammenhängt, wann er anwendbar ist und wann nicht, und was ihn von verwandten Alltagsbegriffen unterscheidet?)
- G3. Rechnen (arithmetisches Operieren)
- G4. Operieren mit mathematischen Termen und Kalkülen
- G5. Interpretation von Diagrammen (gemeint sind hier nicht veranschaulichende Skizzen oder geometrische Konstruktionen, sondern Koordinatendarstellungen oder statistische Diagramme)
- G6. Textverständnis (Müssen komplizierte oder ungewöhnliche sprachliche Formulierungen oder ein längerer Text gelesen und verstanden werden?)
- G7. Bildnerisches Vorstellen
- G8. Problemlöseprozesse

Für Mathematikaufgaben gibt es vier ergänzende Dimensionen:

- M1. Formalisierung (Beschreibung von Zusammenhängen, die in Form von Messreihen, Texten oder Graphen dargestellt sind, in algebraischer Form, das heisst durch Terme, Gleichungen oder Gleichungssysteme)
- M2. Interpretation von Anwendungssituationen (Ist ein Bezug auf aussermathematische Kontexte wesentlich? Muss der Bearbeiter diesen Anwendungskontext verstehen und deuten? Muss er dazu aussermathematische Kenntnisse einbringen?)
- M3. Umstrukturierung/Flexibilität (Dieses Anforderungsmerkmal wird als bedeutungsvoll angesehen, wenn die Lösung der Aufgabe davon abhängt, eine verbal beschriebene Situation oder eine Visualisierung in einer bestimmten, nicht offensichtlichen Weise neu zu „sehen".)
- M4. Prinzipien mathematischen Denkens (Hier geht es quasi um „metamathematisches" Denken. Muss der Bearbeiter beispielsweise zwischen Begriffen, Definitionen, formalsprachlichen Benennungen und Interpretationen (Beispielen) unterscheiden? Hat er die Verallgemeinerbarkeit von Aussagen oder die Gültigkeit eines mathematischen Modells zu überprüfen? Muss er verstanden haben, dass und warum mathematische Aussagen bewiesen werden müssen?)

Es hat sich gezeigt, dass die bei EVAMAR II entwickelten Testaufgaben sehr genau in dieses System passen. Im Gegensatz dazu erwiesen sich die Kategorien M1 – M4 als zu abstrakt und zu umfassend. Die inhaltlichen Gebiete stimmen ebenfalls sehr gut mit den bei EVAMAR II gefundenen Inhalten überein (siehe Abbildung III.16).

3.3 Das Endmodell „Kompetenzraster Mathematik" für EVAMAR II

Während in anderen Projekten wie HarmoS Testaufgaben ausschliesslich auf der Grundlage von Kompetenzmodellen entwickelt werden, sollte der Kompetenzraster in EVAMAR II nur ein grundlegendes Element für die Aufgabenentwicklung sein. Die Aufgaben müssen sich ja auf Inhalte (Sinneinheiten) beziehen, die in Studienunterlagen (Skripten, Bücher etc.) von schweizerischen Universitäten gefunden wurden (siehe Abschnitt 1). Der Kompetenzraster in Mathematik wurde deshalb nicht nur aufgrund der Analyse bestehender Kompetenzmodelle und einer darauf aufbauenden „theoretischen Modellierung" entworfen, seine Konstruktion und ständige Erweiterung erfolgte vielmehr gleichzeitig mit der Entwicklung der Testaufgaben, die sich auf die in Teilprojekt A ermittelten Sinneinheiten in Mathematik beziehen. Die inhaltliche Dimension wurde auf die codierten Sinneinheiten ausgerichtet, und die Dimension der mathematischen Fähigkeiten erst nach einer ersten Serie von Testaufgaben endgültig festgelegt.

Kompetenzen wie Argumentieren, Mathematisieren, Modellieren oder Problemlösen stehen nicht im Zentrum des Interesses von EVAMAR II, da untersucht werden soll, inwieweit Maturandinnen und Maturanden Studientexte von Schweizerischen Universitäten im Hinblick auf mathematische Inhalte verstehen können. Aus diesem Grund konnten die beschriebenen Kompetenzmodelle der KMK sowie von HarmoS und PISA, die diese Fähigkei-

ten mit zugrunde legen, bezüglich der Fähigkeitsdimension nicht übernommen werden. Zudem sollen mit dem Kompetenzraster für EVAMAR II Anforderungen an das Wissen und Können in Mathematik an der Schnittstelle Gymnasium–Universität modelliert werden können. Eine Adaption der für die unteren Klassenstufen konzipierten Kompetenzmodelle erwies sich als zu schwierig, so dass sich die Modelle von KMK, HarmoS und PISA auch wegen der Inhalte als nicht geeignet erwiesen.

Unter den analysierten Kompetenzmodellen ist das Modell bei TIMSS/III (Baumert et al., 2000; siehe Abschnitt 3.2.5) das einzige, welches Kompetenzen von Gymnasiastinnen und Gymnasiasten zum Ende ihrer Schulzeit beschreibt. Interessant für EVAMAR II sind in diesem Modell insbesondere die für Aufgaben entwickelten Kategorien G1 – G7. Es hat sich gezeigt, dass die bei EVAMAR II entwickelten Testaufgaben sehr genau in dieses System passen. Im Gegensatz dazu erwiesen sich die Kategorien M1 – M4 als zu abstrakt und zu umfassend. Die inhaltlichen Gebiete von TIMSS/III stimmen dagegen wiederum sehr gut mit den bei EVAMAR II gefundenen Inhalten überein (siehe Abbildung III.16). Das Modell wurde deshalb als Grundlage für den Kompetenzraster bestimmt.

Dabei erwies es sich bei der inhaltlichen Dimension als zweckmässig, statt der fünf Themen des TIMSS-Modells „Zahlen, Gleichungen und Funktionen“, „Analysis“, „Geometrie und analytische Geometrie“, „Wahrscheinlichkeit und Statistik“ sowie „Aussagenlogik und Beweise“ nur vier inhaltliche Themen zu unterscheiden. Für das Sachgebiet „Aussagenlogik, Beweise“ standen zu wenige Sinneinheiten aus Teilprojekt A zur Verfügung. Daneben fanden sich in den Studienunterlagen Sinneinheiten zu Mengenlehre oder elementarer Algebra. Als Ergebnis wurde deshalb für die Verwendung in EVAMAR II die inhaltliche Haupt-Dimension auf die Hauptgebiete „Analysis“, „Geometrie“, „Stochastik“ und „Vermischtes“ festgelegt. Das letztgenannte Gebiet fasst alles zusammen, was nicht zu den drei anderen Hauptgebieten gehört. Die Unterscheidung in die vier genannten Hauptgebiete ist sowohl für mathematische Inhalte des Gymnasiums wie auch der Hochschule passend. Dies war auch im Hinblick auf Teilprojekt D1 (Analyse von Maturaprüfungen) von Bedeutung. Zudem ergeben sich bei nur vier Gebieten mehr Aufgaben pro Zelle, was für die statistische Auswertung des Tests vorteilhaft ist. Nach Durchsicht der Sinneinheiten aus Teilprojekt A1 und der Entwicklung der Testaufgaben wurde jedes der vier Hauptgebiete zu seiner genaueren inhaltlichen Beschreibung in drei weitere Teilgebiete unterteilt.

Somit ergibt sich für die Inhaltsdimension des für EVAMAR II konstruierten Kompetenzrasters die folgende Gliederung:

- Analysis:
  - Funktionen
  - Infinitesimalrechnung (Differenzial- und Integralrechnung)
  - Folgen und Reihen
- Geometrie:
  - elementare Geometrie
  - Trigonometrie
  - Vektorrechnung
- Stochastik:
  - Wahrscheinlichkeitsrechnung
  - Kombinatorik
  - Statistik
- Vermischtes:
  - Mengenlehre, Diskrete Mathematik
  - Logik
  - elementare Algebra

Die Fähigkeitsdimension des Rasters wurde ebenfalls erst nach der Entwicklung einer ersten Serie von Testaufgaben konkretisiert. Es zeigte sich, dass wieder das bei TIMSS/III entwickelte Kategoriensystem sehr gut passt. Es handelt sich um die folgenden Kategorien (vgl. Baumert et al., 2000):

- Kenntnisse von Definitionen und Sätzen: Es genügt, in den Lehrtexten z. B. den Binomischen Lehrsatz oder den Satz des Pythagoras zu kennen oder zu erkennen.
- Qualitatives Verständnis mathematischer Begriffe: Dazu gehören Aufgaben, bei denen die blosse Kenntnis einer Definition oder eines mathematischen Sachverhaltes nicht ausreicht.
- Rechnen: Dieser Kategorie sind Aufgaben zugeordnet, bei denen der Umgang mit Zahlen ausreicht. Variablen, Unbekannte oder Terme mit Variablen kommen nicht vor.
- Operieren mit mathematischen Termen und Kalkülen: In diese Kategorie gehören Aufgaben, bei denen Terme vorkommen, die nicht nur Zahlen enthalten.
- Interpretieren von Diagrammen: Hier geht es um das Interpretieren von Repräsentationen statistischer Daten wie auch um mathematische Darstellungen wie Funktionsgraphen oder geometrische Zeichnungen.
- Textverständnis: Einige der Aufgaben stellen besondere Anforderungen an das Textverständnis. Aufgaben in dieser Kategorie sind oft zusätzlich einer weiteren Kategorie zugeordnet.
- Bildliches Vorstellen: Das Lösen von Aufgaben dieser Kategorie erfordert Vorstellungsvermögen oder ein Zurechtfinden im dreidimensionalen Raum.

Für EVAMAR II ergibt sich so der in Abbildung III.17 dargestellte Kompetenzraster.

Abb. III.17: Kompetenzraster Mathematik für EVAMAR II

	ANALYSIS			GEOMETRIE			STOCHASTIK			VERMISCHTES		
	Funktionen	Infinitesimal-rechnung	Folgen und Reihen	Trigono-metrie	(analytische) Geometrie	Vektorrechnung	Wahr-scheinlich-keitsrech-nung	Kombinatorik	Statistik	Mengenlehre, Diskrete Ma-thematik	Logik	elementare Algebra
Kenntnisse von Definitionen und Sätzen												
Qualitatives Verständnis math. Begriffe												
Rechnen												
Operieren mit math. Termen und Kalkülen												
Interpretieren von Diagram-men												
Textver-ständnis												
Bildliches Vorstellen												

4 Der Kompetenzraster in Biologie

4.1 Grundlagen

Auch der Kompetenzraster Biologie stützt sich auf die Analyse bereits bestehender Kompetenzbeschreibungen für dieses Fach. Mithilfe dieses Werkzeugs sollten im weiteren Verlauf des Projekts EVAMAR II ebenfalls die Items des Tests in Biologie (Teilprojekt C) auf inhaltlicher Ebene generiert und die Aufgaben aus Maturitätsprüfungen Biologie (Teilprojekt D1) auf inhaltlicher Ebene sowie allenfalls auch auf der Ebene der kognitiven Anforderungsbereiche kategorisiert werden. Im Hinblick auf den Verwendungszweck war dabei einschränkend zu beachten, dass im Teilprojekt C wegen methodischer Beschränkungen und aus Ressourcengründen keine Leistungstests im Bereich bestimmter Handlungskompetenzen wie z. B. Erkenntnisgewinnung, Kommunikation und Bewertung durchgeführt werden konnten. Da ein ideales Kompetenzmodell umfassende Handlungskompetenzen beinhaltet (Klieme, 2003; Weinert, 2001), sollte in Teilprojekt C auch kein umfassender Kompetenztest, sondern eher ein Wissenstest durchgeführt werden. Es ist deshalb besser, im Folgenden von einem „Kompetenzraster Fachwissen Biologie“ zu sprechen.

Der folgende Überblick (Abschnitt 4.2) portraitiert zunächst Einteilungen der Biologie in Teilgebiete, wie sie im In- und Ausland bereits vorgeschlagen wurden. Daran schliesst ein Überblick über bestehende Dimensionierungen entlang kognitiver Anforderungsbereiche biologischer Aufgabenstellungen an. Die Überlegungen münden schliesslich in einen zweidimensionalen Kompetenzraster für das Fachwissen Biologie (Abschnitt 4.3).

4.2 Bestehende Kompetenzraster und -modelle für Biologie

Als auf Schweizer Bundesebene gefasste Anhaltspunkte für Kompetenzen im Bereich Fachwissen auf der Sekundarstufe II könnten die Richtlinien 2003 – 2006 (Gültigkeit verlängert bis 2008) der Schweizerischen Maturitätsprüfung, Bereich Naturwissenschaften, herangezogen werden (SBF, 2008). Es handelt sich dabei um die einzigen spezifisch für die Schweiz gefassten Anhaltspunkte für Kompetenzen im Bereich Naturwissenschaften auf der Sekundarstufe II. Darin sind für das Grundlagenfach Biologie „Bildungsziele“ und ein „Programm“ beschrieben. Sie haben den Charakter von Lernzielen, wie sie in der Schweiz typischerweise jede Fachschaft jeweils für ihre eigene Schule formuliert und als Grundlage für ihre Maturaprüfungen verwendet. Weil aber im Projekt EVAMAR II die Maturitätsprüfungen selbst Untersuchungsgegenstand sind (Teilprojekt D1), können diese Richtlinien von vornherein nicht zur Modellierung des Kompetenzrasters in Biologie verwendet werden, und sie werden deshalb im folgenden Abschnitt nicht weiter analysiert.

Es bietet sich im Weiteren der Blick auf ständig gepflegte, speziell auf wissenschaftliche Inhalte ausgerichtete Klassifikationssysteme an, deren im deutschsprachigen Raum bekanntestes die Regensburger Verbundklassifikation (RVK) sein dürfte.

4.2.1 Die Regensburger Verbundklassifikation (RVK)

Es zeigt sich schnell, dass solche Klassifikationen an die Bedürfnisse der Sacherschliessung in Bibliotheken angepasst und nicht primär nach didaktisch-pädagogischen Kriterien ausgerichtet sind. Für die Bibliotheken sind auch Teilbereiche relevant wie „Referateorgane und Zeitschriften“ oder „Bibliographien und Nachschlagewerke“, „Geschichte und Didaktik“ oder „Lehrbücher der Biologie“, welche natürlich nicht zum Schulstoff gehören.

Wenn man diese zusätzlichen Bereiche ausklammert, legt die RVK folgende Einteilung der Biologie in Teilgebiete vor:

Cytologie; Mikrobiologie; Biotechnologie; Immunologie; Genetik; Evolution; Ökologie; Spezielle Botanik (Systematik); Morphologie und Anatomie der Pflanzen; Pflanzenphysiologie; Spezielle (taxonomische) Zoologie; Verhal-

tensforschung und Tierpsychologie; Anthropologie; Morphologie, Anatomie und Physiologie allgemein, Anatomie, Physiologie und Biochemie einzelner Organe und Organsysteme von Mensch und Tier; Ernährung, Verdauung, Stoffwechsel, Energiehaushalt, Fortpflanzung, Entwicklungsphysiologie von Mensch und Tier, Angewandte Physiologie.

Wie sich weiter unten zeigen wird, lässt sich diese Kategorisierung relativ leicht in die Klassifikationen aus dem schulpädagogischen Bereich einordnen (siehe Abbildung III.21, unterste Zeile). Eine Ausnahme bildet das Teilgebiet „Anthropologie" mit seiner Teilung in die eher materialistisch-physische Richtung („Welche Phylogenese durchlief der Mensch"?) und die eher idealistisch-pragmatische Richtung („Was macht den Menschen zum Menschen?"). Während sich die materialistisch-physische Richtung eindeutig dem Teilgebiet der Evolution zuordnen lässt, müsste für die idealistisch-pragmatische Richtung tatsächlich ein zusätzliches Teilgebiet definiert werden.

Die Betrachtung des RVK bringt als erstes Zwischenresultat, dass der hier entwickelte Kompetenzraster Fachwissen Biologie offenbar sehr umfassend ist und in seiner Breite nur gerade Inhalte der idealistisch-pragmatischen Richtung der Anthropologie nicht aufnehmen kann. Diese wären aber vor dem Hintergrund des untersuchten Konstrukts der „Allgemeinen Studierfähigkeit" vermutlich eher relevant geworden, wenn das Studienfach Philosophie untersucht worden wäre. In den Ergebnissen aus Teilprojekt A lassen sich keine Hinweise finden, dass Inhalte der idealistisch-pragmatischen Richtung der Anthropologie bei der Untersuchung der Studierfähigkeit für die untersuchten Studienfächer, insbesondere auch nicht für das Studienfach Biologie, relevant wären.

Der Rest des Kapitels konzentriert sich nun auf die Eignungsanalyse von Klassifikationen, die im Umfeld der Schulpädagogik entstanden sind. Für die inhaltliche Ebene dürften zwei Betrachtungen ergiebig sein:

1. die Inhaltsanalyse einiger der weitestverbreiteten deutschsprachigen Standardlehrwerke der Biologie auf der Gymnasialstufe (Stufe SII) (Kap. 4.2.2) und zweier amerikanischer Standardwerke auf der Stufe des Studienbeginns (Kap. 4.2.3),
2. die EPA Biologie und Bildungsstandards in Deutschland (Kap. 4.2.4).

Für die Ebene der kognitiven Anforderungsbereiche sind interessant:

1. der Dimensionierungsvorschlag der „Scientific Literacy" aus dem englischsprachigen Raum (Kap. 4.2.6),
2. die EPA Biologie und Bildungsstandards in Deutschland (Kap. 4.2.7 und 4.2.8).

4.2.2 Vergleich der Kapiteleinteilung deutschsprachiger Standardlehrbücher auf der gymnasialen Stufe

Aufgrund ihrer Ausrichtung an spezifisch didaktisch-pädagogischen Zielen und ihrem meist langjährigen Wachsen an zahlreichen Überarbeitungen kann man von einem Standardlehrbuch eine besonders durchdachte und ausgereifte sowie für den schulpädagogischen Anwendungsbereich im Rahmen von EVAMAR II besonders relevante Einteilung des Wissensbereichs erwarten (siehe die Kriterien in Abschnitt 1). Besonders aufschlussreich für die Einteilung der Biologie in Wissensgebiete ist die Kapiteleinteilung der Werke.

Abbildung III.18 zeigt einen Vergleich der Kapiteleinteilung von drei Schülerbänden für den Biologieunterricht, die nur für den Einsatz auf der Stufe SII bestimmt sind und deshalb vornehmlich am Schluss der gymnasialen Ausbildung eingesetzt werden. Diese Werke kommen in der Schweiz schon seit Jahrzehnten regelmässig zum Einsatz. Das Lehrwerk „Linder Biologie" (2005) dürfte den höchsten Bekanntheitsgrad haben. Es ist sofort ein hohes Mass der Übereinstimmung zu erkennen, das natürlich auch als Ausdruck einer wenig hinterfragten Konvention interpretiert werden kann. Wesentlich wahrscheinlicher ist aber, dass die Gleichartigkeit eine sich inhaltlich aufdrängende Konvergenz in der „Evolution" aller denkbaren und je vorgeschlagenen Kapiteleinteilungen darstellt. Die grössten Differenzen zeigen sich dort, wo die Bücher für spezifisch wissenschaftstheoretische

oder spezifisch anwendungsorientierte Inhalte eigene Kapitel abspalten. Die grösste Übereinstimmung findet sich bei den Kapiteln zu Cytologie, Stoffwechsel, Immunbiologie, Ökologie und Evolutionsbiologie. Die inhaltlich sich überlappenden Teilgebiete der Genetik, Fortpflanzung und Entwicklungsbiologie werden unterschiedlich gruppiert. Mit Ausnahme des Gesamtbandes „Biologie Oberstufe“ (Cornelsen Verlag) wird der Entwicklungsbiologie ein eigenes Kapitel gewidmet. Ebenso mit Ausnahme von „Biologie Oberstufe“ erhält das Gebiet der Verhaltensbiologie ein eigenes Kapitel. Ob es hinsichtlich der Verwendung als Schulbuch eher vorteilhaft ist, in weiter gefassten Kapiteln die Querbezüge zu betonen, wie es der Cornelsen Verlag tut, oder – wie die anderen Verlage – durch mehr Kapitel den Überblick zu verbessern, kann unterschiedlich bewertet werden. Im Kontext von EVAMAR II liegt die Hauptanwendung des hier zu entwickelnden Kompetenzrasters darin, die Aufgaben der Hauptuntersuchung sowie ausgewählter Maturitätsprüfungen einem einzelnen Themengebiet sicher zuordnen zu können. Da in biologischen Aufgabenstellungen, besonders in Aufgaben mit einem erhöhten kognitiven Aktivierungspotenzial, häufig nah verwandte Inhaltsbereiche gleichzeitig angesprochen werden, ist es leichter, die Zuteilung vorzunehmen, wenn derlei verwandte Inhaltsbereiche bereits zu relativ umfassenden Teilbereichen zusammengefasst sind.

4.2.3 Vergleich der Kapiteleinteilung von Standardlehrbüchern für Anfänger des Biologiestudiums

Die beiden Übersichtswerke für das Biologiegrundstudium (Campbell & Reece, 2006 und Purves et al., 2008) zeigen eine recht hohe Übereinstimmung untereinander, aber eine eher geringe Kongruenz hinsichtlich der Kapiteleinteilung der Lehrbücher auf Stufe SII (Abb. III.19). Augenfällig ist, dass der Themenbereich „Stoff- und Energiewechsel“ bei beiden Lehrbüchern zusammen mit dem Themenbereich „Kommunikation“ zwischen „Zellen“, „neuronale Informationsverarbeitung“ und „Wahrnehmung“ in einem gemeinsamen Kapitel behandelt wird. Diese damit ausserordentlich umfassenden Kapitel verweben sämtliche Gesichtspunkte, welche zum Verständnis des Funktionierens eines ganzen Organismus (bzw. eines ganzen Organs) gehören. Die Kapiteleinteilung findet nicht entlang von Basiskonzepten der Biologie statt, sondern entlang einer Trennlinie der Systematik, nämlich derjenigen zwischen Pflanzen und Tieren. Die restlichen Inhalte, welche entweder die eindeutig suborganismische Ebene (Chemie des Lebens, einzelne Zelle, Genetik) oder die eindeutig superorganismische Ebene (Ökologie oder Evolutionstheorie) behandeln, werden in eigenen Kapiteln dargestellt.

Abb. III.18: Vergleich der Kapiteleinteilung von drei deutschsprachigen Standardlehrwerken der Biologie auf der Gymnasialstufe (Stufe SII)

Linder Biologie (2005): Gesamtband SII, 22. Auflage. Schroedel, Braunschweig.	Erkenntniswege der Biologie	Cytologie	Stoffwechsel und Energiehaushalt	Genetik	Entwicklungsbiologie	Immunbiologie	Ökologie	Neurobiologie	Hormone	Verhaltensbiologie	Evolution	Baupläne der Lebewesen	--
Biologie heute entdecken (2004). Allgemeine Ausgabe für SII. Schroedel, Braunschweig.	--	Zellbiologie	Stoffwechsel	Genetik	Fortpflanzung und Entwicklung	Immunbiologie	Ökologie	Nerven-, Sinnes- und Hormonphysiologie		Verhaltensbiologie	Evolutionsbiologie	Das System der Lebewesen	Angewandte Biologie
Biologie Oberstufe (2001). Gesamtband. Cornelsen, Berlin.	Biologie - die Wissenschaft vom Leben	Zellbiologie	Stoffwechsel	Vererbung, Fortpflanzung und Entwicklung		Immunbiologie	Ökologie	Informationsverarbeitung, Regelung und Verhalten			Evolution	--	--

Abb. III.19: Vergleich der Kapiteleinteilung zweier übersetzter amerikanischer Standardlehrwerke der Biologie für Studienanfänger (Tertiärstufe)

Campbell, N.A. & Reece, J.B. (2005): Biology, 7th edition. Pearson Education; Upper Saddle River.	The Chemistry of Life	The Cell	Genetics		--	Mechanisms of Evolution	The Evolutionary History of Biological Diversity	Plant Form and Function	Animal Form and Function	Ecology
Campbell, N.A. & Reece, J.B. (2006): Biology, 6. Auflage Pearson Education, München.	Die Chemie des Lebens	Die Zelle	Die Gene		--	Die Mechanismen der Evolution	Die Stammesgeschichte der biologischen Diversität	Form und Funktion der Pflanzen	Form und Funktion der Tiere	Ökologie und Verhalten
Purves, W.K. et al. (2008): Life - The Science of Biology, 8th edition. Sinauer Associates, Sunderland.	The Science and Building Blocks of Life	Cells and Energy	Heredity and the Genome	The Genome in Action	--	The Patterns and Processes of Evolution	The Evolution of Diversity	Flowering Plants: Form and Function	Animals: Form and Function	Ecology
Purves, W.K. et al. (2006): Biologie, 7. Auflage. Spektrum Akademischer Verlag, München.	--	Die Zelle	Information und Vererbung		Entwicklung	Evolutionsprozesse	Die Evolution der biologischen Vielfalt	Die Biologie der Blütenpflanzen	Die Physiologie der Tiere	Ökologie und Biogeographie

4.2.4 Die EPA Biologie und Bildungsstandards in Deutschland

Im Folgenden werden wichtige Texte immer wieder mit den gängigen Kurzformeln zitiert. Abbildung III.20 zeigt die präzisen Bezeichnungen der Dokumente.

Abb. III.20: Kurzformeln für die nachfolgend besprochenen Dokumente

Kurzformel	Kontext
Bildungsstandards KMK Biologie	Bildungsstandards Biologie für den Mittleren Schulabschluss gemäss Beschluss der Kultusministerkonferenz vom 16.12.2004 (KMK, 2005)
EPA Biologie	„Einheitliche Prüfungsanforderungen in der Abiturprüfung Biologie" (Beschluss der Kultusministerkonferenz vom 01.12.1989 i.d.F. vom 05.02.2004 (KMK, 2004)
Stellungnahme des MNU	Briefliche Stellungnahme des Deutschen Vereins zur Förderung des mathematischen und naturwissenschaftlichen Unterrichts e.V. (MNU) zu den Entwürfen der Bildungsstandards für den Mittleren Schulabschluss in Biologie, Chemie und Physik (Stand 30.08. 2004)

Derzeit wird am Leibniz-Institut für die Pädagogik der Naturwissenschaften in Kiel (IPN) im Rahmen des Projekts „Biologie im Kontext" (bik) im Auftrag des Bundesministeriums für Bildung und Forschung (BMBF) ein Kompetenzmodell Biologie für die Stufe SI entwickelt, dessen Kompetenzstufen in der Dimension Fachwissen mit Hilfe von Testfragen empirisch gefunden werden sollen. Die Testfragen werden derzeit (August 2008) konstruiert und in einer nächsten Projektphase an einer Population auf der Stufe SI evaluiert (Ph. Schmiemann, Univ. Duisburg, pers. Mitteilung). Die entstehende Dimensionierung der Dimension „Fachwissen" wird sich mit grösster Wahrscheinlichkeit an die Vorarbeiten der Bildungsstandards KMK Biologie (2005) anlehnen, welche sich ebenfalls auf die Stufe SI beziehen (vgl. Kapitel 4.2.4.2).

4.2.4.1 EPA Biologie

Die EPA Biologie arbeiten mit Bezug auf den Abschluss auf Sekundarstufe II mit zwei hier relevanten „Perspektiven". Es sind dies die Perspektiven der „Themenbereiche" und der „Basiskonzepte" (die dritte Perspektive nennt sich „Reflexionselemente zum Menschenbild"; sie ist hier nicht relevant, siehe Kap. 4.2.1). Da es sich bei diesen „Perspektiven" um unterschiedliche Blickwinkel auf denselben Gegenstand handelt, sind deren jeweilige Inhalte natürlich untereinander relativ eng verflochten.

Inhaltliche Perspektive „Themenbereiche"

Die erste Perspektive, die für eine Einteilung der Biologie in Teilgebiete relevant ist, ist diejenige der Themenbereiche (KMK, 2004, Kap. 1.2.1. Themenbereiche, S. 7f):

Funktionszusammenhänge und deren molekulare Grundlagen

Themen aus der Physiologie, Zellbiologie, Genetik

Von makroskopisch sichtbaren über mikroskopische zu molekularen Zusammenhängen: In einer immer tiefer gehenden Analyse der Ursache-Wirkungsbeziehungen werden Denkprozesse auf verschiedenen Ebenen und hieraus abgeleitete Einsichten in komplexe Zusammenhänge bearbeitet.

Zum Themenbereich Funktionszusammenhänge gehören:

- Bau und Funktion von Zellen, Geweben und Organen; funktionsbezogene Differenzierungen
- Kommunikation zwischen Zellen
- Neuronale Informationsverarbeitung, Wahrnehmung
- Grundlagen der molekularen Genetik
- Stoff- und Energiewechsel: Notwendigkeit und Wege der Energieumwandlung, Assimilation und Dissimilation im Zusammenhang von zellulären Strukturen und Organismus
- Molekulare Steuerung des Stoffwechsels: Enzymatik
- Anwendungen moderner biologischer Erkenntnisse und Methoden, z. B. Gentechnologie, Reproduktionsbiologie, Biotechnologie

Vernetzte Systeme - Ökologie und Nachhaltigkeit

Die Ökologie thematisiert die Wechselwirkungen zwischen Organismus und Umwelt sowie

die Beziehungsgefüge zwischen den Organisationsebenen Population und Ökosystem. Dazu gehört auch das Verhältnis von Mensch und Natur.

Zum Themenbereich Ökologie und Nachhaltigkeit gehören:

- Untersuchung und Analyse eines Ökosystems, z. B. Bestandsuntersuchungen, Messverfahren
- Ökologische Faktoren, Biotop und Biozönose
- Artenvielfalt und Populationsentwicklungen
- Stoffkreisläufe und Energiefluss
- Menschen und Ökosysteme: Beeinflussung von Ökosystemen und deren Folgen, Konsequenzen für den Menschen und die Menschheit

Entwicklungsprozesse - Evolution und Zukunftsfragen

In der Evolution spiegeln sich die Vielfalt der Lebewesen und deren Wechselwirkungen wider.

Betrachtungen zur Evolution beleuchten die Entstehung und das Werden des Lebens als stammesgeschichtlichen Prozess und vermitteln die Einsicht, dass wir Menschen Teil der Evolution sind. Dazu gehören auch Fragen zur zukünftigen Entwicklung des Lebens auf der Erde.

Der Themenbereich Evolution und Zukunftsfragen beinhaltet:

- Stammesgeschichte, Verwandtschaftsbeziehungen, Systematik
- Entstehung der Formen und Arten, Variabilität und Einnischung
- Angepasstheit, auch soziobiologische Fragestellungen
- Evolutionstheorien, Bewertung ihrer Aussagekraft
- Herkunft und Zukunft des Menschen

Inhaltliche Perspektive „Basiskonzepte“
Als Basiskonzepte werden genannt (KMK, 2004, Kap. 1.2.2. Basiskonzepte, S. 8):

a) Struktur und Funktion

Lebewesen und Lebensvorgänge sind an Strukturen gebunden; es gibt einen Zusammenhang zwischen Struktur und Funktion.

Dieses Basiskonzept hilft z. B. beim Verständnis des Baus von Biomolekülen, der Funktion der Enzyme, der Organe und der Ökosysteme.

b) Reproduktion
Lebewesen sind fähig zur Reproduktion; damit verbunden ist die Weitergabe von Erbinformationen.

Dieses Basiskonzept hilft z. B. beim Verständnis der identischen Replikation der DNS, der Viren, der Mitose und der geschlechtlichen Fortpflanzung.

c) Kompartimentierung
Lebende Systeme zeigen abgegrenzte Reaktionsräume.

Dieses Basiskonzept hilft z. B. beim Verständnis der Zellorganellen, der Organe und der Biosphäre.

d) Steuerung und Regelung
Lebende Systeme halten bestimmte Zustände durch Regulation aufrecht und reagieren auf Veränderungen.

Dieses Basiskonzept hilft z. B. beim Verständnis der Proteinbiosynthese, der hormonellen Regulation und der Populationsentwicklung.

e) Stoff- und Energieumwandlung
Lebewesen sind offene Systeme; sie sind gebunden an Stoff- und Energieumwandlungen.

Dieses Basiskonzept hilft z. B. beim Verständnis der Photosynthese, der Ernährung und der Stoffkreisläufe.

f) Information und Kommunikation
Lebewesen nehmen Informationen auf, speichern und verarbeiten sie und kommunizieren.

Dieses Basiskonzept hilft z. B. beim Verständnis der Verschlüsselung von Information auf der Ebene der Makromoleküle, der Erregungsleitung, des Lernens und des Territorialverhaltens.

g) Variabilität und Angepasstheit
Lebewesen sind bezüglich Bau und Funktion an ihre Umwelt angepasst. Angepasstheit wird durch Variabilität ermöglicht. Grundlage der Variabilität bei Lebewesen sind Mutation, Rekombination und Modifikation.

Dieses Basiskonzept hilft z. B. beim Verständnis der Sichelzellanämie, der ökologischen Nische und der Artbildung.

h) Geschichte und Verwandtschaft
Ähnlichkeit und Vielfalt von Lebewesen sind das Ergebnis stammesgeschichtlicher Entwicklungsprozesse.

Dieses Basiskonzept hilft z. B. beim Verständnis der Entstehung des Lebens, homologer Organe und der Herkunft des Menschen.

Da die „Themenbereiche“ und „Basiskonzepte“ unterschiedliche Perspektiven auf ein- und denselben Gegenstand darstellen, wäre es denkbar, sie tabellarisch „gekreuzt“, d. h. in zweidimensionaler Auffächerung darzustellen. Eine derartig feinkammerige Aufteilung biologischer Inhalte ist aber hier nicht angestrebt. Wir haben deshalb versucht, den Themenbereichen diejenigen Basiskonzepte zuzuordnen, welche im jeweiligen Themenbereich besonders oft auftauchen (siehe Abb. III.21).

4.2.4.2 Bildungsstandards KMK Biologie

Bedauerlicherweise ohne explizite Bezüge auf die EPA Biologie oder eine andere Quelle und ohne eine ausführliche Begründung schlagen die Autoren der **Bildungsstandards KMK Biologie** drei Basiskonzepte für den Abschluss auf Sekundarstufe I vor (KMK, 2005, Kap. 2.1, S. 8ff.):

a) System
b) Struktur und Funktion
c) Entwicklung

Diese Reduktion der Basiskonzepte bringt neben den EPA Biologie für EVAMAR II keinen zusätzlichen Nutzen und zwar aus den folgenden beiden Gründen:

1. Stufenargument: **Die EPA Biologie** sind deshalb so wertvoll, weil sie sich explizit auf die Sekundarstufe II beziehen; die **Bildungsstandards KMK Biologie** betreffen die Sekundarstufe I.

2. „Verschlimmbesserungs-Argument“: Hier folgen die Autoren der **Stellungnahme des MNU** zu der vorgenommenen Reduktion der Basiskonzepte:

   „Die Reduzierung der inhaltlichen Dimension auf die 3 Basiskonzepte System, Struktur & Funktion, Entwicklung ist für das Fach Biologie zu allgemein und damit nicht angemessen. Insbesondere das ‚Systemkonzept‘ (S.8) ist in dieser Abstraktheit für Schüler der Sekundarstufe I kein geeignetes Konzept, um Kontexte zu analysieren, Inhalte zu strukturieren und zu systematisieren...“ (S.7f.).
   Die Bildungsstandards für den mittleren Bildungsabschluss sind inkompatibel mit den (gut gelungenen) bundeseinheitlichen EPA (Frühjahr 2004):
   Statt acht Basiskonzepten in den EPA wird ein einzelnes Basiskonzept herausgegriffen und werden zwei weitere auf einem deutlich abstrakteren Niveau hinzugefügt.
   Es fehlen wesentliche Inhalte der Biologie wie Evolution/Selektion, Humanbiologie und Regulation.
   Damit wird die Biologie in ihrer Breite nicht abgebildet, die Basis der Allgemeinbildung wird verkleinert, **Biologie wird im Wissensbereich auf die Reproduktion beschränkt,** statt des durch kumulatives und vernetzendes Lernen mit Hilfe von zahlreicheren und konkreteren Basiskonzepten erreichbaren und notwendigen Verstehens der Biologie.
   Ohne die theoretische Grundlegung auf den Evolutions- und Selektionsgedanken gibt es keine Biologie und keinen sinnvollen Biologieunterricht. Dieses Defizit wird im besonderen Maße in den Ausführungen zum Struktur-Funktions-Konzept deutlich: Der Text ist unscharf, berücksichtigt in keiner Weise die biologietypischen Begründungs- resp. Erklärungsweisen und übersieht, dass eine funktionale Erklärung nur mit dem Bezug auf die Evolutions- und Selektionstheorie erfolgen kann.
   Die differenzierteren Konzepte/Erschließungsfelder: Fortpflanzung, Vielfalt, Angepasstheit, Strukturen & Funktionen, Stoff & Energie, Zeit, Ebenen, Regulation, Wechselwirkung, Information, Eigenart des Menschen werden zum einen der inhaltlichen Vielfalt der Biologie besser gerecht, zum anderen können sie auch von den angesprochenen Personen – Biologielehrkräften sowie Schülerinnen der Sekundarstufe I – konkreter nachvollzogen werden (...).“ (Hervorhebungen im Original)

4.2.5 Zwischenergebnis für die inhaltliche Ebene: Zusammenstellung

Abb. III.21: Einteilung der Biologie in Teilgebiete, angelehnt an deutsch- und englischsprachige Standardbiologielehrwerke, an die deutschen EPA Biologie und an die Regensburger Verbundklassifikation

Werk	Stufe	Cytologie/ Anatomie/Stoffwechsel	Informationsverarbeitung-/Verhalten	Immunbiologie	Genetik/ Entwicklungsbiologie	Ökologie	Systematik/-Evolution
Linder Biologie (2005): Gesamtband SII, 22. Auflage. Schroedel, Braunschweig.	SII	• Cytologie • Stoffwechsel und Energiehaushalt	• Neurobiologie • Hormone • Verhalten	• Immunbiologie	• Genetik • Entwicklungsbiologie	• Ökologie	• Baupläne der Lebewesen • Evolution
Biologie Oberstufe (2001). Gesamtband. Cornelsen, Berlin.	SII	• Zellbiologie • Stoffwechsel	• Informationsverarbeitung, Regelung und Verhalten	• Immunbiologie	• Vererbung, Fortpflanzung und Entwicklung	• Ökologie	• Evolution
Biologie heute entdecken (2004). Allgemeine Ausgabe für SII. Schroedel, Braunschweig.	SII	• Zellbiologie • Stoffwechsel	• Nerven-, Sinnes- und Hormonphysiologie • Verhaltensbiologie	• Immunbiologie	• Genetik • Fortpflanzung und Entwicklung • Angewandte Biologie	• Ökologie	• Das System der Lebewesen • Evolutionsbiologie
Campbell, N.A. & Reece, J.B. (2005): Biology, 7th edition. Pearson Education; Upper Saddle River. Campbell, N.A. & Reece, J.B. (2006): Biology, 6. Auflage. Pearson Education, München.	tertiär	• The Chemistry of Life (Die Chemie des Lebens) • The Cell (Die Zelle) • Plant Form and Function (Form und Funktion der Pflanzen) • Animal Form and Function (Form und Funktion der Tiere)			• Genetics (Die Gene)	• Ecology (Ökologie und Verhalten)	• Mechanisms of Evolution (Die Mechanismen der Evolution) • The Evolutionary History of Biological Diversity (Die Stammesgeschichte der biologischen Diversität)
Purves, W.K. et al. (2008): Life - The Science of Biology, 8th edition. Sinauer	tertiär	• The Science and Building Blocks of Life (--) • Cells and Energy (Die Zelle) • Flowering Plants: Form and Function (Die Biologie der Blütenpflanzen)			• Heredity and the Genome (Information und Vererbung)	• Ecology (Ökologie und Biogeographie)	• The Patterns and Processes of Evolution (The Evolution of Diversity)

Werk	Stufe	Cytologie/ Anatomie/Stoffwechsel	Informationsverarbeitung- /Verhalten	Immunbiologie	Genetik/ Entwicklungsbiologie	Ökologie	Systematik/- Evolution
Associates, Sunderland. Purves, W.K. et al. (2006): Biologie, 7. Auflage. Spektrum Akademischer Verlag, München		• Animals: Form and Function (Die Physiologie der Tiere)			• The Genome in Action (Entwicklung)		• Evolutionsprozesse (Die Evolution der biologischen Vielfalt)
KMK (Hrsg.) (2004): Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland: Einheitliche Prüfungsanforderungen in der Abiturprüfung Biologie (Beschluss der Kultusministerkonferenz vom 01.12.1989 i.d.F. vom 05.02.2004).	SII	**A. Funktionszusammenhänge und deren molekulare Grundlagen** **Themen aus der Physiologie, Zellbiologie, Genetik** Von makroskopisch sichtbaren über mikroskopische zu molekularen Zusammenhängen: In einer immer tiefer gehenden Analyse der Ursache-Wirkungsbeziehungen werden Denkprozesse auf verschiedenen Ebenen und hieraus abgeleitete Einsichten in komplexe Zusammenhänge bearbeitet. Zum Themenbereich Funktionszusammenhänge gehören: • Bau und Funktion von Zellen, Geweben und Organen; funktionsbezogene Differenzierungen • Kommunikation zwischen Zellen • Neuronale Informationsverarbeitung, Wahrnehmung • Grundlagen der molekularen Genetik • Stoff- und Energiewechsel: Notwendigkeit und Wege der Energieumwandlung, Assimilation und Dissimilation im Zusammenhang von zellulären Strukturen und Organismus • Molekulare Steuerung des Stoffwechsels: Enzymatik • Anwendungen moderner biologischer Erkenntnisse und Methoden, z. B. Gentechnologie, Reproduktionsbiologie, Biotechnologie • Hinweis: Immunbiologie wird nicht explizit erwähnt.				**B. Vernetzte Systeme - Ökologie und Nachhaltigkeit** Die Ökologie thematisiert die Wechselwirkungen zwischen Organismus und Umwelt sowie die Beziehungsgefüge zwischen den Organisationsebenen Population und Ökosystem. Dazu gehört auch das Verhältnis von Mensch und Natur. Zum Themenbereich Ökologie und Nachhaltigkeit gehören: • Untersuchung und Analyse eines Ökosystems, z. B. Bestandsuntersuchungen, Messverfahren • Ökologische Faktoren, Biotop und Biozönose • Artenvielfalt und Populationsentwicklungen • Stoffkreisläufe und Energiefluss	**C. Entwicklungsprozesse - Evolution und Zukunftsfragen** In der Evolution spiegeln sich die Vielfalt der Lebewesen und deren Wechselwirkungen wider. Betrachtungen zur Evolution beleuchten die Entstehung und das Werden des Lebens als stammesgeschichtlichen Prozess und vermitteln die Einsicht, dass wir Menschen Teil der Evolution sind. Dazu gehören auch Fragen zur zukünftigen Entwicklung des Lebens auf der Erde. Der Themenbereich Evolution und Zukunftsfragen beinhaltet: • Stammesgeschichte, Verwandtschaftsbeziehungen, Systematik • Entstehung der Formen und Arten, Variabilität und Einnischung

Werk	Stufe	Cytologie/ Anatomie/Stoffwechsel	Informationsverarbeitung- /Verhalten	Immunbiologie	Genetik/ Entwicklungsbiologie	Ökologie	Systematik/- Evolution
						• Menschen und Ökosysteme: Beeinflussung von Ökosystemen und deren Folgen, Konsequenzen für den Menschen und die Menschheit.	• Angepasstheit, auch soziobiologische Fragestellungen • Evolutionstheorien, Bewertung ihrer Aussagekraft • Herkunft und Zukunft des Menschen
		Teilbereich 1 **Bau und Funktion von Zellen, Geweben und Organen; funktionsbezogene Differenzierungen** **Zugeordnetes Basiskonzept Kompartimentierung** Lebende Systeme zeigen abgegrenzte Reaktionsräume. **Zugeordnetes Basiskonzept Struktur und Funktion** Lebewesen und Lebensvorgänge sind an Strukturen gebunden; es gibt einen Zusammenhang zwischen Struktur und Funktion. **Teilbereich 4** **Stoff- und Energiewechsel: Notwendigkeit und Wege der Energieumwandlung, Assimilation und Dissimilation im Zusammenhang von zellulären Strukturen und Organismus,**	**Teilbereich 2** **Kommunikation zwischen Zellen,** **neuronale Informationsverarbeitung, Wahrnehmung** **Zugeordnetes Basiskonzept Information und Kommunikation** Lebewesen nehmen Informationen auf, speichern und verarbeiten sie und kommunizieren. **Zugeordnetes Basiskonzept Steuerung und Regelung** Lebende Systeme halten bestimmte Zustände durch Regulation aufrecht und reagieren auf Veränderungen. .	explizit nicht erwähnt (!)	**Teilbereich 3** **Grundlagen der molekularen Genetik,** **Anwendungen moderner biologischer Erkenntnisse und Methoden, z. B. Gentechnologie, Reproduktionsbiologie, Biotechnologie** **Zugeordnetes Basiskonzept Reproduktion** Lebewesen sind fähig zur Reproduktion; damit verbunden ist die Weitergabe von Erbinformationen.	**Teilbereich 5** **Vernetzte Systeme - Ökologie und Nachhaltigkeit** **Zugeordnetes Basiskonzept** **Stoff- und Energieumwandlung** Lebewesen sind offene Systeme; sie sind gebunden an Stoff- und Energieumwandlungen. **Zugeordnetes Basiskonzept Steuerung und Regelung** Lebende Systeme halten bestimmte Zustände durch Regulation aufrecht und reagieren auf Veränderungen.	**Teilbereich 6** **Entwicklungsprozesse - Evolution und Zukunftsfragen** **Zugeordnetes Basiskonzept Variabilität und Angepasstheit** Lebewesen sind bezüglich Bau und Funktion an ihre Umwelt angepasst. Angepasstheit wird durch Variabilität ermöglicht. Grundlage der Variabilität bei Lebewesen sind Mutation, Rekombination und Modifikation. **Zugeordnetes Basiskonzept Geschichte und Verwandtschaft** Ähnlichkeit und Vielfalt von Lebewesen sind das Ergebnis stammesgeschichtlicher Entwicklungsprozesse.

Werk	Stufe	Cytologie/ Anatomie/Stoffwechsel	Informationsverarbeitung- /Verhalten	Immunbiologie	Genetik/ Entwicklungsbiologie	Ökologie	Systematik/- Evolution
		Molekulare Steuerung von Stoffwechsel: Enzymatik **Zugeordnetes Basiskonzept** **Stoff- und Energieumwandlung** Lebewesen sind offene Systeme; sie sind gebunden an Stoff- und Energieumwandlungen.					
Regensburger Verbundklassifikation (RVK), 2008		• Cytologie • Mikrobiologie • Morphologie und Anatomie der Pflanzen • Pflanzenphysiologie • Morphologie, Anatomie und Physiologie allgemein; Anatomie, Physiologie und Biochemie einzelner Organe und Organsysteme von Mensch und Tier • Ernährung, Verdauung, Stoffwechsel, Energiehaushalt, Fortpflanzung, Entwicklungsphysiologie von Mensch und Tier, Angewandte Physiologie	• Informationsverarbeitung nicht explizit erwähnt • Verhaltensforschung und Tierpsychologie	• Immunologie	Entwicklungsbiologie erwähnt in • Ernährung, Verdauung, Stoffwechsel, Energiehaushalt, Fortpflanzung, Entwicklungsphysiologie von Mensch und Tier, Angewandte Physiologie • Genetik • Biotechnologie	• Ökologie	• Evolution Spezielle Botanik (Systematik) • Spezielle (taxonomische) Zoologie • Anthropologie

Im Weiteren werden bestehende und momentan entstehende Kategorisierungen der Dimension „Kognitive Anforderungsbereiche“ beschrieben. Uns sind drei Kategorisierungen der Dimension „Anforderungsbereiche“ bekannt: Die „Scientific Literacy“ von Bybee sowie wiederum die Bildungsstandards KMK und die EPA Biologie aus Deutschland. In der Schweiz gibt es für diese Dimension auf Stufe SI und SII keine vergleichbaren Ansätze.

4.2.6 Scientific Literacy

Aus dem englischsprachigen Raum stammt der Dimensionierungsvorschlag der „Scientific Literacy“. Bybee (1997) unterscheidet fünf Stufen:

Abb. III.22: Kategorien naturwissenschaftlicher Bildung (Scientific Literacy) nach Bybee (1997). Übersetzung nach Gräber, W. et al. (Hrsg.) (2002): Scientific Literacy. Leske und Budrich, VS Verlag für Sozialwissenschaften/GWV Fachverlage GmbH, Wiesbaden. S. 21 – 43.

Kategorie	Merkmale
Multidimensionale naturwissenschaftliche Bildung (Multidimensional Scientific Literacy)	Naturwissenschaft wird als kulturelle Leistung erkannt, es werden innerhalb und zwischen den Disziplinen wie auch zu gesellschaftlichen Problemfeldern Verknüpfungen hergestellt. Verständnis der philosophischen und historischen Dimension der Disziplinen.
Konzeptuelle und prozedurale naturwissenschaftliche Bildung (Conceptual Scientific Literacy)	Verständnis davon, wie sich die einzelnen Komponenten zum Ganzen der Disziplin und zu den Methoden und Prozessen des Forschers verhalten (Beispiel: Evolution als Konzept bringt Genetik, Stoffwechsel, Struktur und Funktion zusammen).
Funktionale naturwissenschaftliche Bildung (Functional Scientific Literacy)	Texte mit einfachem naturwissenschaftlichem Vokabular können gelesen und geschrieben werden; Wissen besteht aus einer Liste erinnerbarer Definitionen; Erkennen einfacher Zusammenhänge, aber nur im eng umgrenzten Kontext.
Nominale naturwissenschaftliche Bildung (Nominal Scientific Literacy)	Begriffe werden mit Wissensgebieten oder Stoffen assoziiert; Ansammlung von Aufgenommenem ohne wirkliches Verstehen (naive Theorien, Fehlkonzepte).
Ungebildetsein (Illiteracy)	Kognitive Unfähigkeit, Fragestellungen zu Naturwissenschaft und Technologie zu verstehen, oder diese den Naturwissenschaften zuzuordnen.

4.2.7 Bildungsstandards KMK

Die Skala der Bildungsstandards KMK nimmt keinen Bezug auf das Konzept der Scientific Literacy. Sie scheint sich stark auf die EPA Biologie zu stützen und unterscheidet die drei folgenden Anforderungsbereiche (vgl. KMK, 2005, Kapitel 4.1, S. 16):

Anforderungsbereich I: Sachverhalte, Methoden und Fertigkeiten reproduzieren.

Dieses Anspruchsniveau umfasst die Wiedergabe von Fachwissen und die Wiederverwendung von Methoden und Fertigkeiten.

Anforderungsbereich II: Sachverhalte, Methoden und Fertigkeiten in neuem Zusammenhang benutzen.
Dieses Niveau umfasst die Bearbeitung grundlegender bekannter Sachverhalte in neuen Kontexten, wobei das zugrunde liegende Fachwissen bzw. die Kompetenzen auch in anderen thematischen Zusammenhängen erworben sein können.

Anforderungsbereich III: Sachverhalte neu erarbeiten und reflektieren sowie Methoden und Fertigkeiten eigenständig anwenden.
Dieses Niveau umfasst die eigenständige Erarbeitung und Reflexion unbekannter Sachverhalte und Probleme auf der Grundlage des Vorwissens. Konzeptwissen und Kompetenzen werden u. a. genutzt für eigene Erklärungen, Untersuchungen, Modellbildungen oder Stellungnahmen.

4.2.8 EPA Biologie

Die EPA Biologie unterscheiden ebenfalls drei Bereiche (KMK, 2004, Kap. 2.1, S. 13), die nicht als in anderen Arbeiten verankert dargestellt werden:

Die Anforderungsbereiche unterscheiden sich vor allem im Grad der Selbstständigkeit bei der Bearbeitung der Aufgaben sowie im Grad der Komplexität der gedanklichen Verarbeitungsprozesse, sie stellen damit eine Abstufung in Bezug auf den Anspruch der Aufgabe dar.

Der Anforderungsbereich I umfasst

- die Verfügbarkeit von Daten, Fakten, Regeln, Formeln, mathematischen Sätzen usw. aus einem begrenzten Gebiet im gelernten Zusammenhang.
- die Beschreibung und Verwendung erlernter und eingeübter Arbeitstechniken und Verfahrensweisen in einem begrenzten Gebiet in einem wiederholenden Zusammenhang.

Im Fach Biologie gehören dazu

- die Reproduktion von Basiswissen (Kenntnisse von Fakten, Zusammenhängen und Methoden)
- die Nutzung bekannter Methoden und Modellvorstellungen in vergleichbaren Beispielen
- die Entnahme von Informationen aus Fachtexten und Umsetzen der Informationen in einfache Schemata (Stammbäume, Flussdiagramme o. Ä.)
- die schriftliche Darstellung von Daten, Tabellen, Diagrammen, Abbildungen mit Hilfe der Fachsprache
- die Beschreibung makroskopischer und mikroskopischer Beobachtungen
- die Beschreibung und Protokollierung von Experimenten
- das Experimentieren nach Anleitung und die Erstellung mikroskopischer Präparate
- die sachgerechte Benutzung bekannter Software

Der Anforderungsbereich II umfasst

- selbstständiges Auswählen, Anordnen, Verarbeiten und Darstellen bekannter Sachverhalte unter vorgegebenen Gesichtspunkten in einem durch Übung bekannten Zusammenhang
- selbstständiges Übertragen des Gelernten auf vergleichbare neuartige Fragestellungen, veränderte Sachzusammenhänge oder abgewandelte Verfahrensweisen.

Im Fach Biologie gehören dazu

- die Anwendung der Basiskonzepte in neuartigen Zusammenhängen
- die Übertragung und Anpassung von Modellvorstellungen

- die sachgerechte, eigenständig strukturierte und aufgabenbezogene Darstellung komplexer biologischer Abläufe im Zusammenhang einer Aufgabenstellung
- die Auswahl bekannter Daten, Fakten und Methoden zur Herstellung neuer Zusammenhänge
- die gezielte Entnahme von Informationen aus vielschichtigen Materialien oder einer wissenschaftlichen Veröffentlichung unter einem vorgegebenen Aspekt
- die abstrahierende Darstellung biologischer Phänomene wie die zeichnerische Darstellung und Interpretation eines nicht bekannten mikroskopischen Präparats
- die Anwendung bekannter Experimente und Untersuchungsmethoden in neuartigen Zusammenhängen
- die Auswertung von unbekannten Untersuchungsergebnissen unter bekannten Aspekten
- die Beurteilung und Bewertung eines bekannten biologischen Sachverhalts
- die Unterscheidung von Alltagsvorstellungen und wissenschaftlichen Erkenntnissen.

Der Anforderungsbereich III umfasst

- planmässiges und kreatives Bearbeiten vielschichtiger Problemstellungen mit dem Ziel, selbstständig zu Lösungen, Deutungen, Wertungen und Folgerungen zu gelangen
- bewusstes und selbstständiges Auswählen und Anpassen geeigneter erlernter Methoden und Verfahren in neuartigen Situationen.

Im Fach Biologie gehören dazu

- die Entwicklung eines eigenständigen Zugangs zu einem biologischen Phänomen, z. B. die Planung eines geeigneten Experimentes oder Gedankenexperimentes
- die selbstständige, zusammenhängende Verarbeitung verschiedener Materialien unter einer selbstständig entwickelten Fragestellung
- die Entwicklung eines komplexen gedanklichen Modells bzw. die eigenständige Modifizierung einer bestehenden Modellvorstellung
- die Entwicklung fundierter Hypothesen auf der Basis verschiedener Fakten, experimenteller Ergebnisse, Materialien und Modelle
- die Reflexion biologischer Sachverhalte in Bezug auf das Menschenbild
- die materialbezogene und differenzierte Beurteilung und Bewertung biologischer Anwendungen
- die Argumentation auf der Basis nicht eindeutiger Rohdaten: Aufbereitung der Daten, Fehleranalyse und Herstellung von Zusammenhängen
- die kritische Reflexion biologischer Fachbegriffe vor dem Hintergrund komplexer und widersprüchlicher Informationen und Beobachtungen

4.3 Beschreibung des EVAMAR-Kompetenzrasters für Biologie

Abbildung III.23 zeigt diejenigen Komponenten, welche unter Berücksichtigung der in Abschnitt 1 beschriebenen Aspekte der Fokussierung in den Kompetenzraster Fachwissen Biologie eingeflossen sind. Gut ersichtlich stützt es sich bei der inhaltlichen Dimension auf die Analyse der drei deutschsprachigen Standardlehrwerke der Biologie auf der Gymnasialstufe (Stufe SII, siehe Abb. III.18). Diese Werke stehen am ehesten für eine Einteilung des Fachgebiets Biologie, wie sie auch im gymnasialen Biologieunterricht verwendet wird. Dabei lassen sich zehn Inhaltsbereiche ermitteln, welche in allen Standardlehrbüchern eigene Kapitel erhalten oder sogar mit mehreren Kapiteln vertreten sind.

Für die Verwendung im Rahmen der Teilprojekte C und D1 ist diese zehnteilige Dimensionierung aber noch zu fein. Der Cornelsen Verlag (2001) hat sich in seinem Unterrichtswerk „Biologie Oberstufe“ für eine Zusammenlegung von sehr stark überlappenden Kapiteln entschieden. So werden die Genetik und die Entwicklungsbiologie zusammengenommen und die drei stark vernetzten Gebiete „Neurobiologie“, „Hormone“ und „Verhalten“ zu einem einzigen Teilgebiet zusammengefasst. Die resultierende siebenteilige Dimensionierung ist aber immer noch zu fein für die Anwendung im Kontext von EVAMAR II. Bei einer gegebenen Testfrage (Teilprojekt C) oder Maturaaufgabe (Teilprojekt D1) ist es oft schwierig zu entscheiden, ob sie eher zur Cytologie oder eher zur Stoffwechselphysiologie gehört. Deshalb sind diese Gebiete zusammen mit Anatomie zu einem grossen Teilgebiet „Cytologie/Anatomie/Physiologie“ zusammengeführt, wobei die Anatomie auch die Lehre von allen äusseren und inneren Formen und Strukturen umfasst, selbst wenn diese sehr klein sind (Mikroanatomie). Wie in Abb. III.21 ersichtlich, entspricht die für EVAMAR II gewählte inhaltliche Einteilung der Biologie fast vollständig derjenigen des Lehrbuches „Biologie Oberstufe“.

Auch bei der Dimensionierung der kognitiven Anforderungsbereiche soll vorrangig das Kriterium der stufengerechten Ausrichtung zur Anwendung kommen. Die Einteilung nach Bybee (1997) ist bewusst stufenunabhängig formuliert und zeigt das gesamte Spektrum von vollständigem Ungebildetsein bis zur ganzheitlichen Schau auf das System der Naturwissenschaften und ist damit weniger geeignet. Von den beiden verbleibenden Einteilungen beziehen sich nur die EPA explizit auf die Stufe SII. Der Mittlere Schulabschluss in Deutschland entspricht in der Schweiz dem Ende der obligatorischen Schulzeit, also dem Abschluss der Stufe SI, sodass die Bildungsstandards KMK Biologie nicht in Frage kommen. Zusätzlich sind die drei Anforderungsbereiche in den EPA Biologie präziser ausformuliert als in den Bildungsstandards KMK Biologie. Die dreistufige Einteilung ist genügend grob, um eine zweifelsfreie Einteilung der Items der Testaufgaben in Teilprojekt C und der in Teilprojekt D1 analysierten Matura-Prüfungsaufgaben zu ermöglichen.

Abb. III.23: Der Kompetenzraster Fachwissen Biologie für die Verwendung in EVAMAR II

		Inhaltsbereich (angelehnt an deutschsprachige Standardbiologielehrwerke SII)					
		Cytologie/ Anatomie/- Stoffwechsel	Informations- verarbei- tung/Verhalten	Immunbiologie	Genetik/- Entwicklungs- biologie	Ökologie	Systematik/- Evolution
Kognitiver Anforderungsbereich (EPA Biologie)	I						
	II						
	III						

IV Projektteil C: Entwicklung und Durchführung von Tests

1 Allgemeines und Übersicht

1.1 Zielsetzungen und Auswahl der Erhebungsgegenstände

Im Rahmen des Teilprojekts C wurden Erhebungen bei Maturandinnen und Maturanden durchgeführt. Sie setzen sich aus Tests in drei Fächern, einer Zusatzbefragung und einem fachübergreifenden kognitiven Fähigkeitstest in der Deutschschweiz zusammen.

Die Ergebnisse dieser Messungen können nicht direkt-ursächlich auf die Maturitätsbildung zurückgeführt werden – das wäre nur über Längsschnittstudien möglich. Sie geben aber den Stand von Wissen, Können, Kompetenzen und weiteren Eigenschaften und Merkmalen kurz vor der Matura wieder. Dies ermöglicht den Vergleich mit einem potenziell maximal möglichen Stand an der Schnittstelle Gymnasium–Universität und die zielgerichtete Suche nach Erklärungen für Abweichungen und deren Struktur sowie Empfehlungen für das künftige Gymnasium.

In der Zusatzbefragung wurde in Ergänzung zum Projekt EVAMAR I auch nach der Selbsteinschätzung einiger weiterer Kompetenzen im Rahmen der Studierfähigkeit gefragt. Das erwies sich als sinnvoll, weil die in den Teilprojekten A1 und A3 ermittelten Daten und Erkenntnisse in ihrer Aussagekraft hinsichtlich der Abbildung des Konstrukts Studierfähigkeit beschränkt waren und weil auch die Ergebnisse der Befragung von Dozierenden an Universitäten in Teilprojekt A4 (siehe Kapitel II) diese Ergänzung nahelegten. In der Deutschschweiz gestaltete sich dieser Fragebogen aufgrund der dort verfügbaren Befragungszeit etwas umfassender.

Der überfachliche Fähigkeitstest (UFT) hat sich aufgrund der Erkenntnisse aus dem Teilprojekt A2 aufgedrängt (siehe Abschnitt 4 in Kapitel II). Er enthält Items, die in anderen Untersuchungen gut mit dem Studienerfolg in naturwissenschaftlichen Fächern korrelierten. Der UFT setzt sich aus validierten Items aus vier Dimensionen des Eignungstests für das Medizinstudium (EMS) zusammen, die einen besonders hohen Vorhersagewert für Prüfungserfolge in universitären Fachstudien besitzen. Dieser Zusatztest wurde nur in der Deutschschweiz eingesetzt, weil gemäss der Einschätzung der Projektleitung und der Deutung verschiedener Signale die Beanspruchung zusätzlicher Test-Zeit in der Romandie und im Tessin nicht mehr zumutbar war. Deshalb wurde dort auf diese Ergänzung verzichtet, zumal sie im ursprünglichen Konzept nicht vorgesehen war und auch nicht in den Bulletins kommuniziert wurde.

Insgesamt sollten diese beiden zusätzlichen Erhebungen weitere Aussagen ermöglichen, erstens im Hinblick auf erweiterte Aspekte der Studierfähigkeit und zweitens über das Verhältnis von standardisierten Wissenstests und allgemeineren Kompetenzen.

Schliesslich wurden auch die effektiv erzielten Maturanoten in den Testfächern erhoben.

1.2 Präzisierungen zur Bestimmung der Testbereiche

1.2.1 Zur Prognosevalidität der Tests im Hinblick auf die Studierfähigkeit

Mit den Tests sollten Aspekte der Studierfähigkeit gemessen werden. Im Folgenden werden einige weitergehende Überlegungen zu deren Prognosevalidität angestellt[22].

Da sich die tatsächliche Studierfähigkeit erst im Verlauf eines Studiums herausstellt, können zum Zeitpunkt der Matura grundsätzlich nur Konstrukte getestet werden, die aufgrund bisheriger empirischer Untersuchungen

[22] Hinweise zur Prognosevalidität von Maturanoten finden sich im Abschnitt 1 von Kapitel I.

einen mehr oder weniger grossen Vorhersagewert für die Studierfähigkeit gezeigt haben. In der relevanten, fast ausschliesslich englischsprachig publizierten Literatur finden sich für die „Allgemeine Studierfähigkeit“ Konstrukte wie „academic achievement“, „academic aptitude“, „academic performance“ oder „academic success“. Je nach Schwerpunkt untersuchen die Arbeiten (sich überschneidende) Aspekte

- der allgemeinen kognitiven Leistungsfähigkeit
- bereits erworbenen Wissens
- der Lesefähigkeit
- der Persönlichkeit oder
- der Motivation

Die nachfolgende Literaturübersicht soll in aller Kürze einen Überblick über die bisherigen Ergebnisse der Forschung liefern.

Allgemeine kognitive Leistungsfähigkeit

Dass zwischen den kognitiven Fähigkeiten einer Person und ihrer akademischen Leistungsfähigkeit ein Zusammenhang besteht, ist plausibel und bestens dokumentiert (siehe die zitierten Studien in Chamorro-Premuzic, 2007). Dennoch erklärt die Punktzahl in IQ-Tests selten mehr als 25% der Varianz der akademischen Leistungsfähigkeit (Chamorro-Premuzic & Furnham, 2004; O’Connor & Paunonen, 2007; Neisser et al., 1996), in der Regel gar weit weniger. Dies wird darauf zurückgeführt, dass Studierende bereits eine homogenere kognitive Leistungsfähigkeit aufweisen als die Gesamtbevölkerung (Furnham, Chamorro-Premuzic & McDougall, 2003). Interessant ist im Zusammenhang mit motivationalen Faktoren (siehe unter e), dass die Studierstrategie des Deep Processing mit dem im Wonderlic Personnel Test (nach Wonderlic, 1992) gemessenen IQ korreliert (r = .32) und dieser wiederum mit universitären Prüfungsnoten (r = .24) (Chamorro-Premuzic & Furnham, 2008).

In EVAMAR II werden die relativ wissensunabhängigen kognitiven Fähigkeiten mit dem überfachlichen Fähigkeitstest (UFT) gemessen.

Bereits erworbenes Wissen

Beim Hören und Lesen neuer Information fördert Vorwissen die Tiefe des Verständnisses sowie die später erinnerte Menge und dessen fachliche Korrektheit. Es ist empirisch gut belegt, dass die Qualität bereits erworbenen Wissens massgeblich darüber bestimmt, wie feinmaschig und gleichzeitig umfassend der Textinhalt mit dem Vorwissen verknüpft werden kann, inwiefern diese Verknüpfung fachlich adäquat erfolgt und wie gut die neuen Inhalte erinnert werden (z. B. Bartlett, 1932; Ausubel, 1968; Bloom, 1976; Chiesi, Spilich & Voss, 1979; Wilson & Anderson, 1986; Recht & Leslie, 1988; Alexander et al., 1989; Dochy, 1992; Tobias, 1994; Kim & Van Dusen, 1998; Thompson & Zamboanga, 2004; Shapiro, 2004; Baumert et al., 2007; Hailikari, Nevgi & Komulainen, 2008; Ozuru, Dempsey, McNamara, in press). Dochy, Segers & Buehl (1999) zeigen in ihrer Metastudie, dass 95% von 183 ausgewählten Publikationen einen fördernden Effekt von bereichsspezifischem Vorwissen auf Lernleistungen nachweisen konnten. Insbesondere wenn der Lehrtext wenig inneren Zusammenhalt (geringe Kohäsion) aufweist, können Leserinnen und Leser mit einem umfassenden Vorwissen während des Lesens leichter erkenntnisfördernde Schlüsse über den lokalen und globalen Zusammenhang zwischen den einzelnen Sätzen ziehen. Dies erlaubt ihnen, beim Lesen eine kohärentere Repräsentation des Textes aufrechtzuerhalten und schliesslich im Gedächtnis abzulegen (McNamara & Kintsch, 1996; McNamara, 2001). Ein geringes Mass an Kohäsion des Textes liegt vor,

- wenn zahlreiche für das Verständnis entscheidende Begriffe, Konzepte und Theorien als Hintergrundwissen vorausgesetzt werden und die einzelnen Aussagen des Textes nicht explizit mit diesen Begriffen, Konzepten und Theorien verwoben sind,
- wenn der Text wenig Redundanz in der Argumentation aufweist,
- wenn der Text Gedankensprünge und fachliche Ungenauigkeiten enthält,
- wenn der Text nicht durch Überschriften, Hervorhebungen, Textkästen usw. hilfreich gegliedert ist (Best, Rowe, Ozuru & McNamara, 2005).

Ein gut strukturiertes und reiches Vorwissen ist deshalb besonders im universitären Kontext wichtig, da hier häufig Lehrtexte mit geringer Kohäsion anzutreffen sind, sei es, weil dieses Vorwissen (aus ausbildungsökonomischen Gründen) vorausgesetzt werden muss, oder aber, weil die Texte durchaus noch Verbesserungspotenzial haben (z. B. Beck, McKeown, Sinatra & Loxterman, 1991; Britton, Gulgoz & Glynn, 1993). Vorwissen hat jedoch nicht immer einen positiven Effekt auf die Tiefe des Textverständnisses, da es oft von Misskonzeptionen durchsetzt ist und mit den in einem Lehrtext entwickelten Gedanken interferieren kann (z. B. West & Pines, 1985; Diakidoy, 1999), so dass es erfolgreiches Lernen behindert (Alexander & Judy, 1988; Vosniadou & Brewer, 1994; Kendeou & van den Broek, 2007; Hailikari, Nevgi & Komulainen, 2008). Interessanterweise scheint es nicht so zu sein, dass Misskonzeptionen den Leseprozess verlangsamen und auf diese Weise den Lernprozess behindern, sondern dass Leser mit Misskonzeptionen häufig relativ geringe Ansprüche an die angestrebte Verständnistiefe beziehungsweise Kohärenzstandards haben (siehe die Erklärung mit dem Landscape Model unten) (van den Broek, Risden & Husebye-Hartmann, 1995; Kendeou & van den Broek, 2007) und deshalb die Widersprüche zwischen ihrem Vorwissen und dem Lehrtext nicht mit zusätzlichem Aufwand zu ergründen versuchen (z.B. mittels Zurückspringen im Text, weiteren Gedächtnisaktivierungen, Konsultation externer Hilfsmittel wie Lexika oder Internet). Die Mängel im Lernerfolg offenbaren sich dann erst bei einem Leistungstest (Kendeou & van den Broek, 2007).

Die Elaboriertheit des Vorwissens inklusive (latenter) Misskonzeptionen wird in EVAMAR II in Wissens- und Könnenstests in den Fächern Mathematik und Biologie getestet.

Lesefähigkeiten (Lesefertigkeit, Lesestrategien und -techniken)

Ein hohes Mass an Lesefertigkeit sowie an entsprechenden Lesetechniken kann bis zu einem gewissen Grad mangelndes Vorwissen kompensieren (Garner, 1987; Voss & Silfies, 1996; O'Reilly & McNamara, 2007). Lesefertigkeit definiert sich bei wissenschaftlichen Texten vor allem über die Fähigkeit, zwischen einzelnen Sätzen und ganzen Textabschnitten Verbindungen zu erkennen (Fähigkeit des inference making, siehe die Referenzen in O'Reilly & McNamara, 2007). Die dazugehörigen Lesetechniken umfassen Fähigkeiten wie sich zuerst einen Überblick verschaffen, in eigenen Worten die Kernaussagen wiederholen oder zusammenfassen, schwierige Passagen mehrmals lesen, im Text vor- und zurückspringen, laut lesen, den Stand des Verständnisses ständig kritisch reflektieren, visuelle Markierungstechniken einsetzen, Handbücher verwenden usw. (Mokhtari & Reichard, 2002).

In EVAMAR II werden Sprachfertigkeiten in Erstsprache getestet, wobei sich der Hauptteil der Items auf die Lesefertigkeit bezieht.

Vorwissen und Lesefähigkeiten im Kontext

Welche Verstehensprozesse während des Lesens stattfinden und wie Vorwissen die Gedächtnisrepräsentation eines Textes beeinflusst, wird seit Jahren theoretisch und empirisch untersucht. Die derzeit meistbeachtete Synthese der Erkenntnisse aus zahlreichen Einzelstudien ist das Landscape Model (van den Broek, Risden, Fletcher & Thurlow, 1996; van den Broek, Young, Tzeng & Linderholm, 1999; van den Broek, Rapp & Kendeou, 2005), das Ähnlichkeiten zum Construction-Integration Model (Kintsch, 1988, 1998) und anderen Zweistufenmodellen aufweist (z.B. Long, Seely & Oppy, 1996; Richards & Singer, 2001). Inzwischen sind die Grundannahmen des Landscape Models in ein Computermodell implementiert. Die aus einzelnen Simulationsläufen gewonnenen Voraussagen können so mit empirischen Daten verglichen werden und ermöglichten eine beeindruckende Feinvalidierung des Modells (van den Broek, Virtue, Everson, Tzeng & Sung, 2002; van den Broek, Kendeou, Sung & Chen, 2003; Tzeng, van den Broek, Kendeou & Lee, 2005).

Die eigentliche Leistung des Landscape Models besteht darin, gedächtnisbasierte und konstruktionistische Prozesse zu verbinden und ihre Wechselwirkung bis auf die Ebene von einzelnen Lesezyklen hinunter zu modellieren.

Gemäss dem gedächtnisbasierten Ansatz der Textverarbeitung löst die gerade gelesene Information (zusammen mit der Information, die ebenfalls gerade im Arbeitsgedächtnis aktiviert ist) die selektive Aktivierung weiterer

Information aus, die im Langzeitgedächtnis damit assoziiert ist. Je elaborierter die Assoziation zwischen der gerade gelesenen Information und einer Information im Langzeitgedächtnis, desto leichter, schneller und verlässlicher kann die soeben gelesene Information jene im Langzeitgedächtnis aktivieren. Auch kann die Aktivierung von Information innerhalb der gerade entstehenden episodischen Repräsentation des bis dahin gelesenen Textes (dem Diskursmodell) erfolgen. Diese ist ja noch nicht im Langzeitgedächtnis abgelegt, befindet sich aber auch nicht im Arbeitsgedächtnis. Gedächtnisbasierte Prozesse werden in dem Sinne als „kostenlos" angesehen, als sie passiv sind und autonom ablaufen, das heisst nicht dem Einfluss einer bewussten Strategie des Lesers unterliegen.

Gemäss dem konstruktionistischen Ansatz der Textverarbeitung versuchen Leser beim Aufnehmen eines Textes aktiv, implizite und explizite Ziele oder Standards in Bezug auf die Konstruktion von Bedeutung beziehungsweise das Erlangen einer gewissen Verständnistiefe zu erreichen. Um diese so genannten Kohärenzstandards (standards of coherence) zu erreichen, können Leser gezielt Informationen abrufen aus dem bereits gelesenen Text (z. B. noch einmal zurückspringen), aus der sich entwickelnden Repräsentation des zu lesenden Textes im Gedächtnis oder aus dem Langzeitgedächtnis. Wenn sich die auf diese Weise zusammengetragenen Informationen widersprechen, werden bei genügend hohen Kohärenzstandards weitere konstruktionistische Prozesse angeregt, um den Widerspruch zu lösen (van den Broek & Kendeou, 2008). Nicht-adäquates Vorwissen und Misskonzeptionen können bei hohen Kohärenzstandards also den Lesefluss deutlich behindern. Die Kohärenzstandards in Bezug auf die gewünschte Verständnistiefe hängen davon ab, ob ein Text zur Unterhaltung oder zu Studienzwecken gelesen wird. Die Höhe der Kohärenzstandards bestimmt wesentlich den Anteil konstruktionistischer Prozesse beim Lesen.

Das Landscape Model verbindet nun diese beiden Ansätze. Es geht davon aus, dass beide Prozesse dynamisch interagieren, wechselseitig voneinander profitieren, aber auch im Widerstreit stehen können. Diese Interaktion stellen sich Van den Broek und Mitarbeiter folgendermassen vor: Der Leser schreitet in einer Folge von Lesezyklen durch den Text, wobei ein Lesezyklus ungefähr einem Satz oder allgemein einer Informationseinheit entspricht. Pro Lesezyklus bewertet der Leser das Resultat einer durch ein Textstück gerade ausgelösten passiven Wissensaktivierung entsprechend den von ihm gewählten Kohärenzstandards. Falls die aktivierte Information nicht ausreicht, um den Text in der gewünschten Tiefe zu verstehen, werden weitere konstruktionistische Anstrengungen unternommen (Zurückspringen im Text, weitere Gedächtnisaktivierungen, Konsultation externer Hilfsmittel wie Lexika oder Internet). Wenn die Kohärenzstandards sehr niedrig angesetzt sind, unterbleiben konstruktionistische Anstrengungen auch dann, wenn die aktivierte Information nicht ausreicht, um eine kohärente Textrepäsentation entstehen zu lassen.

Faktoren der Persönlichkeit

Eine kürzlich veröffentlichte Metastudie kommt nach der Analyse von rund 20 ausgewählten Studien zum Schluss, dass die Persönlichkeitsdimension der Gewissenhaftigkeit (Conscientiousness) den weitaus stärksten und konsistentesten Prädiktor für die akademische Leistungsfähigkeit (academic performance) darstellt (mittleres r = 0.24, 90%-Vertrauensintervall zwischen r = 0.12 und r = 0.36) (O'Conner und Paunonen, 2007). Die Dimension der Conscientiousness ist neben Neuroticism, Extraversion, Openness to Experience und Agreeableness einer der Faktoren des Five-Factor Models der Persönlichkeit (McCrae & Costa, 1997). Unterfaktoren der Gewissenhaftigkeit wie Leistungsstreben (achievement-striving), Selbstdisziplin (self-discipline) und Pflichtbewusstsein (dutifulness) sind noch stärker mit der tatsächlichen akademischen Leistungsfähigkeit korreliert als der übergeordnete Persönlichkeitsfaktor selbst. Der Unterfaktor „Leistungsstreben" beschreibt die Eigenschaft, ehrgeizig, strebsam, fleissig, sorgfältig und ausdauernd zu sein. Der Unterfaktor „Selbstdisziplin" umfasst folgende Merkmale: motiviert sein, Aufgaben zu Ende bringen und standhaft gegenüber Ablenkungen sein. Der Unterfaktor „Pflichtbewusstsein" bezeichnet das subjektiv empfundene Bedürfnis, moralischen Verpflichtungen nachzukommen (O'Conner und Paunonen, 2007).

Gemäss Art. 5 des MAR 95 (1995) sollen die Gymnasien auch die „Willenskraft" fördern, was zumindest als indirekter Hinweis darauf verstanden werden kann, dass der gymnasiale Unterricht diese drei Eigenschaften

auch fördern soll. Im Rahmen von EVAMAR II werden Daten zu Leistungsstreben und zur Selbstdisziplin im Rahmen des Zusatzfragebogens erhoben.

Motivationale Faktoren

Eine gute Übersicht liefert der bisher umfassendste Versuch, empirisch bereits gut belegte motivationale Faktoren der akademischen Leistungsfähigkeit zu einem integrativen konzeptuellen Modell zusammenzuführen durch Fenollar, Roman & Cuestas (2007). Gemäss ihrem zentralen Gedanken werden Strategien der Informationsverarbeitung (oberflächlich vs. tief) und des Energieeinsatzes (effort) von gewissen Leistungszielen beziehungsweise dem subjektiv wahrgenommenen Mass an Selbstwirksamkeit (Self-efficacy) geformt. Diese Leistungsziele haben einerseits direkte Effekte auf die akademische Leistungsfähigkeit und andererseits indirekte Wirkungen auf die Art der Strategien der Informationsverarbeitung. Fenollar, Roman & Cuestas (2007) lassen vier relativ unabhängige Leistungsziele in ihr Modell einfliessen, die sich in der Literatur zunehmend etablieren (Ames & Archer, 1988 und Elliot & Thrash, 2001):

- Mastery Goals: Kompetenzen entwickeln und Wissen vermehren sowie durch angestrengtes Lernen Verständnis und Einsicht in den Lernstoff gewinnen
- Performance-approach Goals: gegen aussen im Wettbewerb erfolgreich sein, andere Studenten übertrumpfen
- Performance avoidance Goals: nicht inkompetent erscheinen, also Inkompetenz verstecken (auf Kosten langfristigen Lernens)
- Work Avoidance Goals: mit einem Minimum an Aufwand die Arbeit erledigen und herausfordernde Aufgaben vermeiden

Das Mass der Selbstwirksamkeit umfasst die subjektiven Überzeugungen, dass man eine Handlung erfolgreich ausführen kann, die zur Erreichung eines Ziels notwendig ist (Bandura 1977).

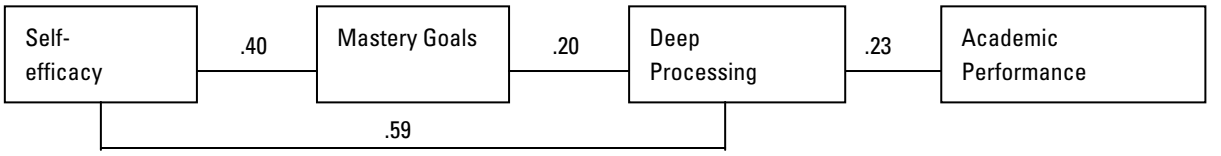
Die Unterscheidung in tiefes beziehungsweise oberflächliches Verarbeiten von Informationen wurde erstmals von Marton & Säljö (1976) vorgeschlagen und später von Tait & Entwistle (1996) in der kognitionspsychologischen Literatur weiter verankert:

Studenten mit einer Strategie des Deep Processing versuchen aktiv, die Bedeutung des Lerninhalts zu verstehen, indem sie Beziehungen zwischen den Konzepten ausarbeiten, Verbindung zwischen dem Vorwissen und dem Lernstoff herstellen und eine kritische Haltung einnehmen.

Studenten mit einer Strategie des Surface Processing fokussieren auf Memorisierungsstrategien, ohne den Versuch, die Bedeutung des Lerninhalts zu verstehen.

Fenollar, Roman & Cuestas (2007) konnten die Hypothese bestätigen, dass Studierstrategien zwischen Leistungszielen und akademischer Leistungsfähigkeit vermitteln. Sie fanden hochsignifikante Korrelationen ($p < 0.01$) zwischen folgenden Parameterpaaren (je verbunden durch eine Linie) (siehe Abbildung IV.1):

Abb. IV.1: Studierstrategien, Leistungsziele und Studienerfolg



Die positiven Zusammenhänge zwischen diesen vier Parametern werden durch zahlreiche weitere Studien belegt (siehe die Referenzen in Fenollar, Roman & Cuestas, 2007; siehe auch die Zusammenhänge im oben beschriebenen Landscape Model des Leseverständnisses).

Selbstverständlich ist es auch hier eine Aufgabe des Gymnasiums, durch guten Unterricht Maturandinnen und Maturanden heranzubilden, die sich zutrauen, auch anspruchsvolle Probleme lösen zu können (Self-efficacy), das

Bedürfnis haben, den Lehrstoff wirklich zu verstehen und nicht nur auswendig zu lernen oder oberflächlich zu reüssieren (Mastery Goals und Deep Processing).

Inwieweit diese Ziele der Studierfähigkeit erreicht werden, wird im Rahmen von EVAMAR II teilweise für b) untersucht.

Die nachfolgende Tabelle IV.1 stellt die wichtigsten Prädiktoren hoher Studierfähigkeit zusammenfassend dar. Sie alle können und sollen im Rahmen der gymnasialen Ausbildung im Sinne von Art. 5 des MAR 1995 gefördert und etabliert werden. Es wird ersichtlich, dass in EVAMAR II eine breite Palette kognitiver Faktoren (allgemeine kognitive Leistungsfähigkeit, Vorwissen und Sprachfähigkeiten) und darüber hinaus Ausschnitte von Persönlichkeitsfaktoren erhoben werden, so dass ein vielschichtiges Bild der relevanten Prädiktoren der allgemeinen Studierfähigkeit entsteht.

Tab. IV.1: Wichtige Prädiktoren hoher Studierfähigkeit

Aspekt	Starke Prädiktoren der allgemeinen Studierfähigkeit	Wichtige zusammenfassende Studien oder Metastudien	Art der Testung in EVAMAR II
allgemeine kognitive Leistungsfähigkeit	▪ Allgemeine kognitive Fähigkeiten / Intelligenz	Chamorro-Premuzic (2007)	Items des überfachlichen Fähigkeitstests (UFT)
bereits erworbenes Wissen	▪ Elaboriertheit des Vorwissens ▪ Zahl und Relevanz von Misskonzeptionen	Dochy, Segers & Buehl (1999)	Tests (Wissen und Können) in Mathematik und Biologie
Lesefähigkeit	▪ Lesefertigkeit ▪ Breite des Repertoires an eingesetzten Lesestrategien und -techniken	O'Reilly & McNamara (2007)	Tests von Sprachfähigkeiten in Erstsprache, insbesondere von Lesefertigkeiten
Faktoren der Persönlichkeit	▪ Leistungsstreben ▪ Selbstdisziplin ▪ Pflichtbewusstsein	O'Conner und Paunonen (2007)	teilweise mittels einzelner Items im Zusatzfragebogen
motivationale Faktoren	▪ Self-efficacy: sich zutrauen, auch anspruchsvolle Probleme lösen zu können ▪ Mastery Goals und Deep Processing: das Bedürfnis haben, den Lehrstoff wirklich zu verstehen und nicht nur auswendig zu lernen oder oberflächlich zu reüssieren	Fenollar, Roman & Cuestas (2007)	teilweise mittels einzelner Items im Zusatzfragebogen

1.2.2 Zur Wahl von Biologie als naturwissenschaftliches Fach

Wie bereits in Kapitel I beschrieben, sollten die Tests gemäss Auftrag nur in den drei Fächern Erstsprache, Mathematik und einem – zum Zeitpunkt der Auftragserteilung noch durch die Steuergruppe zu bestimmenden – naturwissenschaftlichen Fach entwickelt und durchgeführt werden. Die Steuergruppe hat – unter Einbezug erster Ergebnisse aus dem Teilprojekt A – im Dezember 2006 aus den drei möglichen Fächern Biologie, Chemie und Physik das Erstgenannte gewählt. Wichtige Teile der Begründung für diese Wahl sind die folgenden: Präzisiert für die Naturwissenschaften wird im Teilprojekt C die Fähigkeit des Maturanden beziehungsweise der Maturandin gemessen, in einem schriftlichen Wissens- und Könnenstest ohne Hilfsmittel zeigen zu können, dass

sie/er sich an allgemein studiumsrelevante naturwissenschaftliche Fakten, Konzepte und Prozeduren tatsächlich erinnern kann und diese verstanden hat. Welche naturwissenschaftlichen Fakten, Konzepte und Prozeduren als „allgemein studiumsrelevant“ gelten, richtet sich danach, welche naturwissenschaftlichen Fakten, Konzepte und Prozeduren in den Lehr-/Lernmaterialien und den ersten Zwischen- oder Assessmentstufen-Prüfungen des ersten Semesters *aller* untersuchten Studienfächer tatsächlich als bekannt und verstanden vorausgesetzt werden (Ergebnisse aus Teilprojekt A).

Die Frage, ob die tatsächlich geprüften naturwissenschaftlichen Fakten, Konzepte und Prozeduren primär biologischer, chemischer oder physikalischer „Natur“ sein sollten, wurde auf der Basis der folgenden zwei Abwägungen beantwortet:

1. Da letztlich immer das Konstrukt der „allgemeinen Studierfähigkeit“ untersucht werden soll, galt es zu entscheiden, welches der drei Fächer Biologie, Chemie oder Physik bis zur Matura am ehesten solcherlei Fakten, Konzepte und Prozeduren lehrt, die bei Studienbeginn in möglichst vielen der untersuchten Studienfächer vorausgesetzt werden.
2. Aus Überlegungen der maximalen Ausbeute galt es ausserdem zu entscheiden, welches der drei Fächer Biologie, Chemie oder Physik bis zur Matura am ehesten solcherlei Fakten, Konzepte und Prozeduren lehrt, die am wenigsten bereits durch die Messung der mathematischen Fähigkeiten im Mathematiktest erhoben werden.

Die erste Abwägung erbrachte eine klare Antwort zugunsten der Biologie. Im Unterschied zu Physik und Chemie greifen nämlich nicht nur die technisch-naturwissenschaftlichen Wissenschaften, sondern auch die Geistes- und Sozialwissenschaften auf biologische Konzepte zurück. Ein Beispiel dafür ist etwa der Einbezug der evolutionären Sichtweise in die Wirtschaftswissenschaften oder in die Psychologie. Gemäss den Ergebnissen aus Teilprojekt A spielen die Chemie und die Physik in den Geistes- und Sozialwissenschaften ebenfalls eine weniger grosse Rolle. Die Biologie überspannt von allen Naturwissenschaften wohl tatsächlich am leichtesten die „Kulturgrenze“ zwischen den „humanities“ und der „science“ (siehe dazu Snow, 1963).

Die zweite Abwägung erbrachte eine Antwort zuungunsten der Physik. Obwohl sich natürlich die studiumsrelevanten Fakten, Konzepte und Prozeduren aus dem Physikunterricht von den mathematischen unterscheiden, tauchen vor allem gewisse Prozeduren, aber auch einige Konzepte aus der Mathematik bei der Lösung eines typischen Physikproblems immer wieder auf. Vor allem auf der Sekundarstufe II müssen physikalische Probleme oft zwingend zuerst in die mathematische Sprache von Gleichungen übersetzt werden, um dann wiederum mit mathematischen Methoden gelöst zu werden. Für die Verwandtschaft der kognitiven Anforderungen spricht auch die Tatsache, dass gemäss Untersuchungen gymnasiale Schulnoten in Physik den Mathematiknoten ähnlicher als die Biologie-/Chemie-Noten sind (Heller, 2002).

Die erste und zweite Abwägung zusammen sprachen für die Wahl der Biologie. Bei der Chemie sprach zwar die zweite Abwägung nicht wie beim Fach Physik explizit gegen, aber die erste Abwägung auch nicht für die Chemie. Somit fiel die Wahl auf Biologie, und die im Rahmen von EVAMAR II geprüften naturwissenschaftlichen Fakten, Konzepte und Prozeduren sollten primär biologischer „Natur“ sein.

2 Beschreibung der Erhebungsinstrumente

2.1 Vorgehen zur Konstruktion der Tests: Grundsätzliches für alle Tests

Als Grundlage für die Konstruktion von Aufgaben dienten die in der Ergebnisdatenbank der Teilprojekte A1 und A3 gespeicherten Wissens- und Könnenselemente, die entsprechenden konkreten Lehrmaterialien sowie die im Teilprojekt B erstellten Kompetenzraster. Dabei wurde eine möglichst repräsentative Verteilung auf die Kompetenzfelder einerseits und die ermittelten Wissens- und Könnenselemente andererseits angestrebt. Die Ergebnis-

se der Befragung von Dozierenden in Teilprojekt A4 erforderten – mit der bereits erwähnten Ausnahme in Erstsprache – keine Korrektur dieser Grundlage.

Durch die Inhaltsanalysen von Studienmaterialien (Teilprojekt A1) wurde eine Fülle von Fachausdrücken, Klassifikationen, Kategorien, Prinzipien, Theorien und Modellen identifiziert, die als Eingangswissen eingestuft werden müssen, da sie im gesichteten Unterrichtsmaterial nicht mehr erklärt, sondern vorausgesetzt werden.

Jedes Test-Item bezieht sich auf eine Textstelle der Studienmaterialien (Teilprojekt A1). Jedes Item wurde konstruiert mit dem Ziel zu messen, ob eine Maturandin oder ein Maturand eine Sinneinheit in einer gewissen Tiefe verstanden hat. Sinneinheiten können Fachausdrücke, Kategorien, Klassifikationen, Prinzipien bis hin zu einer ganzen Theorie beziehungsweise einem ganzen Modell sein (siehe auch Kapitel II). Entscheidend ist, dass das Verständnis der Sinneinheit nicht im entsprechenden Lehrmaterial vermittelt, sondern vorausgesetzt wird. Die Items erheben also das Ausmass und die Qualität des (vorausgesetzten) Eingangswissens beziehungsweise -verständnisses. Dabei wurde von den folgenden Brückenhypothesen ausgegangen:

1. Je besser das fachliche Eingangswissen, desto besser ist das fachliche Textverständnis (vgl. Abschnitt 1.2.1). Je mehr Items korrekt beantwortet werden können, desto leichter wird im Studium das Verständnis der Lehrmaterialien an denjenigen Stellen fallen, an denen sich der Text auf entsprechende Fachausdrücke, Klassifikationen, Kategorien, Prinzipien, Generalisierungen, Theorien und Modelle bezieht.
2. Je besser das fachliche Textverständnis, desto höher ist die Studierfähigkeit in dem jeweiligen Fach, in dem solche Sinneinheiten verwendet werden.
3. Je besser das fachliche Textverständnis ist und je leichter der Abruf der fachlichen Sinneinheiten im Text eines beliebigen Studienfachs aus dem Langzeitgedächtnis sowie deren Verarbeitung fällt, desto ausgeprägter ist die allgemeine Studierfähigkeit.

Zusätzlich zu den allgemein gültigen Leitlinien für die Konstruktion von Testaufgaben erfordert die Untersuchung der Studierfähigkeit gemessen an den tatsächlichen Studienanforderungen, dass die Items so realistisch wie möglich die konkreten kognitiven Anforderungen widerspiegeln, mit denen die Studentin beziehungsweise der Student im ersten Semester konfrontiert ist. Alle Items beziehen sich auf die Situation des Lesens und Verstehens von Studienunterlagen (Skripte und Bücher). Manchmal sind deshalb die Sinneinheiten im Aufgabenstamm und in den Optionen möglichst genau in ihren tatsächlich vorliegenden Kontext eingebettet. Dies macht die Items in den meisten Fällen schwieriger, als wenn die Sinneinheit in der Aufgabenformulierung für sich allein stehen würde.

Die Items für den überfachlichen Fähigkeitstest wurden durch das Zentrum für Testentwicklung und Diagnostik Fribourg (ZTD) zur Verfügung gestellt.

Um die ganze Breite der Kompetenzraster einbeziehen zu können, wurden insgesamt Testaufgaben entwickelt, die von einem Schüler in 720 Minuten gelöst werden könnten. So viel Zeit stand für die Evaluation nicht zur Verfügung, weshalb nicht jeder Maturandin beziehungsweise jedem Maturand alle Aufgaben vorgelegt werden konnten. Pro Test standen lediglich 45 Minuten zur Verfügung. Für eine zuverlässige Schätzung der Fähigkeiten der Maturandinnen und Maturanden (Personenparameter), aber auch der Schwierigkeit der Aufgaben (Schwierigkeitsparameter), wurde deshalb ein rotierendes Testdesign gewählt, bei dem die Aufgaben untereinander verlinkt sind (Rasch Common Item Equating beziehungsweise Multi-Matrix-Design, bei dem die Testhefte über gleiche Aufgabenblöcke miteinander verknüpft sind, so dass sich das für die einzelne Person ermittelte Wissen und Können nicht nur auf das von ihr bearbeitete Testheft stützen muss) (Wright & Stone, 1979). Für die Biologie mussten dementsprechend sieben Hefte erstellt werden, für die Mathematik sechs, für die Erstsprache Deutsch fünf, für die Erstsprachen Italienisch und Französisch je vier und für die überfachlichen Kompetenzen ebenfalls vier Hefte. Die Testhefte wurden zufällig auf die Maturandinnen und Maturanden verteilt, sodass jede Aufgabe jeweils von ähnlich vielen Maturandinnen und Maturanden bearbeitet wurde. Tabelle IV.2 zeigt die Verteilung der Blöcke auf die Hefte exemplarisch für Biologie.

Tab. IV.2: Die sieben Testhefte in Biologie

	Heft 1	**Heft 2**	**Heft 3**	**Heft 4**	**Heft 5**	**Heft 6**	**Heft 7**
Heft-Hälfte A	BIO 1	BIO 2	BIO 3	BIO 4	BIO 5	BIO 6	BIO 7
Heft-Hälfte B	BIO 2	BIO 3	BIO 4	BIO 5	BIO 6	BIO 7	BIO 1

Bei der Wahl der Aufgabenformen dominierten wie bei vergleichbaren grossen Untersuchungen (TIMSS, PISA) die Kriterien Auswertungsobjektivität und Auswertungsökonomie. Es wurden möglichst wenig Aufgaben zur ausführlichen offenen Beantwortung eingesetzt (am meisten in Erstsprache), weil diese zur Sicherung der Objektivität aufwändige Codierungsarbeiten nach sich ziehen. Damit sich trotzdem die ganze beabsichtigte Breite von Kompetenzen erfassen liess, konnte aber vor allem in Erstsprache nicht auf diesen Aufgabentyp verzichtet werden. Sorgfältige Schulung der Korrektorinnen und Korrektoren und weitere Massnahmen sollten dennoch zu grösstmöglicher Objektivität führen. Die Beschränkung auf die schriftliche Durchführungsform hatte auch eine Einschränkung der testbaren Kompetenzfelder in Erstsprache zur Folge (z. B. Verzicht auf Hör- und Sprechkompetenzen).

2.2 Beschreibung der qualitativen Expertenevaluation und der Pretests (inkl. Verfahren und Überarbeitung der Items)

Die in einer ersten Fassung erstellten Items wurden dem folgenden Qualitätsprüfungs- und Revisionsverfahren unterzogen:

- Beurteilung durch Experten unmittelbar nach der Testkonstruktion und erste Revision
- Durchführung der Tests bei rund 100 bis 180 Maturandinnen und Maturanden von Abschlussklassen an ausgewählten Gymnasien; Überprüfung und allfällige Revision der Items aufgrund der Kriterien Rasch-Modellpassung „Infit“, Trennschärfekoeffizient der Einzel-Items, ausgewogene Distraktorenverteilung und Rasch-konformer Verlauf der „Item Characteristic Curve“ (ICC) (siehe die Definitionen in Abschnitt 2.4)

Die Schulen wurden nach pragmatischen Kriterien ausgewählt (Zugänglichkeit, Bereitschaft zur Mitarbeit). Die Durchführung der Vortests erfolgte Ende Februar/Anfang März 2007.

Die qualitative Beurteilung der Aufgaben in Erstsprache Deutsch erfolgte durch einen Experten der Pädagogischen Hochschule Zürich. Ein erster Vortest mit einer Version noch ohne Items zur Messung der grammatikalischen Kompetenz wurde im Sinne eines Pilotversuches bereits im September 2006 an der Kantonsschule Zug mit 65 Maturandinnen und Maturanden durchgeführt, und die Aufgaben wurden bereits damals ein erstes Mal revidiert. Der zweite Vortest mit allen Items im Anschluss an die Expertenvalidierung erfolgte an den Kantonsschulen Sargans und Frauenfeld mit rund 180 Teilnehmenden. Für die in Französisch und Italienisch übersetzten Items erfolgte je ein weiterer Vortest. Alle Test-Items wurden nach den Pretests in allen drei Sprachregionen (siehe auch Kapitel 1.5.2) einer gemeinsamen weiteren Optimierung unterzogen.

Die qualitative Beurteilung der Aufgaben in Mathematik erfolgte durch Experten der ETHZ. Aufgrund des äusserst knapp bemessenen Zeitbudgets waren aber noch nicht alle Aufgaben für diese erste Etappe der Validierung bereit. Ein Teil der Aufgaben wurde deshalb im Rahmen der qualitativen Validierung der in die französische Sprache übersetzten Version ein weiteres Mal begutachtet. Der Vortest wurde an der Alten Kantonsschule Aarau bei acht Maturaklassen mit 163 Maturandinnen und Maturanden durchgeführt. In der Folge wurden die Testaufgaben sprachlich und inhaltlich überarbeitet und mit neuen Aufgaben ergänzt.

Der Vortest in Biologie erfolgte am Freien Gymnasium Zürich bei 110 Schülerinnen und Schülern.

2.3 Die Übersetzungen in die zwei anderen Landessprachen: Grundsätzliches für alle drei Fachgebiete

Die Übersetzung der Aufgaben in die beiden anderen Landessprachen Italienisch und Französisch ist aus verschiedenen Gründen ein sehr anspruchsvoller Vorgang. Diskutiert wurden zwei Verfahrensvarianten:

- Variante 1: Übersetzung der deutschen Items in die andere Sprache und Rückübersetzung durch eine andere Person in Deutsch. Bei Abweichungen erfolgt eine Korrektur der Übersetzung.
- Variante 2: zweimalige Übersetzung der Items durch zwei Personen, Vergleich der Ergebnisse und Diskussion sowie Anpassungen bei Abweichungen.

Aufgrund der Erfahrungen bei internationalen Untersuchungen wie PISA und TIMSS lässt sich feststellen, dass Variante 1 wegen der häufig entstehenden Abweichungen wesentlich ressourcenintensiver ist. Diese Abweichungen lassen sich trotz des Einsatzes hoch qualifizierter Übersetzender nicht vermeiden. Aus Ressourcengründen wurde deshalb für EVAMAR II grundsätzlich Variante 2 gewählt.

2.4 Vorgehen zur Auswahl der in die Hauptauswertungen einbezogenen Items

In die Hauptauswertungen sollten nur jene Aufgaben aufgenommen werden, die auch aufgrund der Haupterhebung teststatistischen Mindestanforderungen genügen. Dazu gehören die folgenden:

- Übereinstimmung mit der Rasch-Modellierung, ausgedrückt durch den Infit: Er zeigt für jede Aufgabe, wie viele unerwartete Antworten unter der Annahme des Rasch-Modells beobachtet werden. Der Infit hat unter dieser Annahme einen Erwartungswert von 1. Weichen die Fit-Werte statistisch signifikant von 1 ab, dann passt die Aufgabe nicht zum Rasch-Modell. Der Infit sollte zwischen 0.8 und 1.2 liegen.
- Mindesttrennschärfekoeffizient von 0.2 (zur Begründung siehe z.B. Bühner, 2006, S. 140; Eberle, 1986). Unter der Trennschärfe wird die punktbiseriale Korrelation der Aufgabe mit dem Test-Rohwert verstanden. Ein hoher Trennschärfekoeffizient bedeutet beispielsweise, dass die Maturandinnen und Maturanden mit hohen Test-Rohwerten die Aufgabe richtig lösen und solche mit niedrigen Test-Rohwerten die Aufgabe falsch lösen.
- Der Verlauf der „Item Characteristic Curve“ (ICC) der einzelnen Items muss Rasch-konform beziehungsweise stetig sein.
- Die fehlende Personenhomogenität zwischen den Sprachregionen darf nicht durch Übersetzungsmängel verursacht sein: Für die Überprüfung der Übersetzung wurde die differenzielle Item-Funktion („Differential Item Function“) (DIF) berechnet, das heisst, es wurde geprüft, ob eine Wechselwirkung zwischen Testaufgabe und Sprachregion bestand. Eine differenzielle Item-Funktion liegt dann vor, wenn Testaufgaben für Personengruppen gleicher Fähigkeit, aber unterschiedlicher Sprachregionen unterschiedliche Lösungswahrscheinlichkeiten besitzen. Das Problem dabei ist, dass differenzielle Item-Funktionen sowohl durch Übersetzungsschwierigkeiten als auch durch unterschiedliche Lehrpläne bedingt sein können. Die für diese Prüfung angewandten Kriterien waren je nach Fach unterschiedlich streng:
  - In Erstsprache durfte die Differenz zwischen den Sprachregionen Deutsch und Französisch höchstens 0.5 Logits auf der Skala der Item-Schwierigkeiten betragen. Falls diese Grenze überschritten wurde, wurde das Item trotzdem dann nicht ausgeschieden, wenn die Differenz zwischen den grossen Deutschschweizer Regionen Zürich und Bern ebenfalls beachtlich war (ausserhalb des 99%-Signifikanzbereiches). Letzteres dient als Indikator dafür, dass der Unterschied nicht übersetzungsbedingt, sondern auf andere Ursachen zurückzuführen ist. Denn gerade solche Unterschiede sollen mit der Untersuchung sichtbar gemacht werden.
  - In Mathematik musste die Differenz zwischen den Sprachregionen Deutsch und Französisch im 99%-Signifikanzbereich liegen; falls diese Grenze überschritten wurde, erfolgte in den Fällen keine Ausschei-

dung des Items, in denen die Differenz zwischen den grossen Deutschschweizer Regionen Zürich und Bern ebenfalls beachtlich war (ausserhalb des 99%-Signifikanzbereiches). Wie oben konnte man dann von nicht übersetzungsbedingten Ursachen ausgehen. Auch hier sollten mit EVAMAR II gerade solche Unterschiede aufgedeckt werden.

- In Biologie musste die Differenz zwischen den Sprachregionen Deutsch und Französisch vorerst im 99%-Signifikanzbereich liegen. Weil aufgrund der primären Kriterien nur noch 56 Items übrig blieben, wurde das Sprachdifferenzkriterium gelockert und demjenigen bei der Erstsprache wie folgt angepasst: Die Differenz zwischen den Sprachregionen Deutsch und Französisch durfte höchstens 0.5 Logits auf der Skala der Itemschwierigkeiten betragen. Diese Lockerung kann zusätzlich damit begründet werden, dass aufgrund der in Biologie sprachregional getrennt durchgeführten Lehreraus- und -weiterbildung Unterschiede im Schwierigkeitsgrad Unterschiede in den Curricula widerspiegeln und nicht einfach auf Übersetzungsfehler zurückgeführt werden können. Das Differenzkriterium „Deutschschweizer Regionen Zürich und Bern“ erbrachte keine zusätzlich verwendbaren Items. Die Hauptauswertungen wurden in der Folge aufgrund der gelockerten Auswahlvariante gerechnet und dargestellt. Falls sich in den parallel durchgeführten Kontrollrechnungen mit der strengen Variante Unterschiede ergeben sollten, werden diese ebenfalls beschrieben und die Unterschiede analysiert.

Im Ergebnis resultierten als Grundlage für die Hauptauswertungen 124 Items in Erstsprache, 117 Items als Grundlage für die Hauptauswertungen in Mathematik und 56 (strenge Variante) beziehungsweise 91 Items (grosszügigere Variante) für Biologie. Für den überfachlichen Fähigkeitstest konnten alle 42 Items auch in die Hauptauswertungen übernommen werden.

2.5 Der Test in Erstsprache

2.5.1 Ergänzendes zur Testentstehung

2.5.1.1 Besonderes und Beispiel zur Aufgabenkonstruktion

Im Test für Erstsprache ging es nicht um eigentliches Fachverständnis wie in Mathematik und Biologie. Die Aufgaben sollten vielmehr im Sinne von Sprachverständnis die Kompetenzen gemäss Kompetenzraster Erstsprache (siehe Berichtsteil III) erfassen, und zwar anhand der in Teilprojekt A untersuchten universitären Lehrtexte. Damit war von vornherein klar, dass sich die mit den Tests in Erstsprache gemessenen Kompetenzen im Vergleich zu den anderen beiden Fächern am meisten von dem im Rahmen von entsprechenden Maturaprüfungen erfassten Wissen und Können unterscheiden würden.

Die Vorauswahl des konkreten Studienmaterials geschah durch Priorisierung von frühem Studienmaterial, also von Material, das bereits zu Beginn des ersten Semesters gelesen werden muss. Innerhalb der gewählten Fachbücher wurde die erste Priorität möglichst auf die ersten Kapitel gelegt, die innerhalb des angegebenen Vorlesungsablaufes zu bearbeiten waren. Dieses sozusagen chronologische Auswahlverfahren wurde gewählt, um möglichst nahe entlang dem erwarteten Vorwissen zu testen.

Für die Konstruktion von Items im Bereich „Wortschatz“ wurden häufig codierte Sinneinheiten als so genanntes Wortschatz-Wissen im Kontextzusammenhang des entsprechenden Studienmaterials abgefragt, dies meist im Multiple-Choice-Format.

Für die Konstruktion von Items sowohl für das allgemeine und detaillierte Leseverstehen als auch für den Grammatikteil musste der Fokus erweitert werden auf den ganzen Textzusammenhang, in dem die codierte Sinneinheit oder die Sinneinheiten-Häufung lag. Um möglichst geeignete Formate schaffen zu können, wurde bei der konkreten Aufgabenstellung die feste Bindung an die codierte Sinneinheit gelöst. Das bedeutet, dass nicht jede Frage der Tests eine der codierten Sinneinheiten im Wortlaut als richtige Antwort erwartet oder sich kon-

kret auf die codierte Sinneinheit bezieht, sondern sich die Lösung auch im näheren Umfeld einer codierten Sinneinheit befinden kann.

Zur Veranschaulichung soll im Folgenden an einem exemplarischen Beispiel erläutert werden, wie dieses Verfahren „vom Lehrbuch zur konkreten Aufgabe“ anhand der Sinneinheiten aussieht.

In der später folgenden Abbildung IV.4 wird ein Originalauszug aus dem in der Pädagogik als Studienmaterial verwendeten Buch *„Lernen: Zwanzig Szenarien aus dem Alltag“* von Gerhard Steiner (2001) gezeigt, so wie er bearbeitet nach der Codierung und für die Auswahl aussieht. Die codierten Sinneinheiten wurden in Wortlisten tabellarisch erfasst (siehe den Auszug für das erste Kapitel des gewählten Buches in Abbildung IV.2). Da die Fremdwörter (Fw) sowohl dem Fach Deutsch als auch den Herkunftssprachen zugeordnet sein konnten, wurden in solchen Fällen auch die Codierungen der Herkunftssprachen Französisch (F), Italienisch (I), Englisch (E), Lateinisch (L) und Griechisch (Gr) in die Auswahl einbezogen.

Abb. IV.2: Auszug von codierten Sinneinheiten (Fremdwörter)[23]

F-I-E-L-Gr-Faktenwissen **AA:**			
2.5. Buch-/Skripttitel: Steiner, G. (2001). *Lernen. Zwanzig Szenarien aus dem Alltag.* 3., korr. Auflage. Bern: Hans Huber. **Dateiname Kodierung: a1_paed_L_Steiner1_Kp1_CI.rtf**			
F: Verhaltensrepertoires **E:** distress stimulus ? **behavior**istische(n)(2) unconditioned stimulus unconditioned response conditioned stimulus conditioned response ?**Behavior**ist **L:** Labor(3) differenziert Differenzierung visuell komplexer(e/n)(4)	Komplexität Signal(tafel)(2) Organismus(5) Lern**prinzipien** Neben**prinzipien** Prinzip selektiv präzise ? **Kognition**spsychologie Inventar Konfiguration **Signal**system(s)(4) antizipiert ? kognitiv(er)(3) ? kognitive(n/r) Prozesse(6) klassische	Experiment(2) experimentelle(r)(2) elaborierten existenzbedrohender Adaption Antizipation verbal Experimentator Präsentation Korrelaten verbal(er)(4) Verbalisieren Präzision differenziertes Vokabular Lokalisierung präzisen	**Gr:** lern**theoretischen** Theorie(n)(4) Grundmechanismus akustische ? Lern**theoretiker**(4) ? Kognitions**psychologie** Signal**system(s)**(4) psychischen psychologischen **Psychologie**geschichte drastisch Phänomen hirn**physiologischen** Nerven**system** physiologischen

Gemäss dem oben erwähnten chronologischen Verfahren handelt es sich beim Textbeispiel um die erste Seite des ersten Kapitels „Angst vor weissen Kitteln – Klassisches Konditionieren“ (S. 15–27), zu der sinnvollerweise eine Aufgabe konstruiert werden konnte. In dem kurzen einleitenden Kapitel 1.1 und im Kapitel 1.2 zu Pawlows Entdeckung fanden sich zwar bereits Codierungen wie „weisse Kittel“, „Neurophysiologe“, „Theorien“ und „Labor“

[23] Mit der Zahl in Klammern (2) wird bei einer codierten Sinneinheit angegeben, wie häufig sie innerhalb dieses Kapitels gefunden wurde, ohne dass sie erklärt oder definiert wurde. Mit der Hervorhebung in Zusammensetzungen gibt die Codiererin an, welchen Teil der Sinneinheit sie der jeweiligen Herkunftssprache zuweist.

(S. 15), diese traten aber zu wenig gehäuft beziehungsweise in einem wenig geeigneten Kontext auf. In Kapitel 1.3 des Buches findet sich nun im Abschnitt „Angstreaktion als elementare Verhaltensweisen“ (Abb. IV.4) relativ zur Textlänge eine erste Häufung von codierten Sinneinheiten (Unterstreichungen und Codes).

Nachdem der Textinhalt des Kontexts um diese gefundene Häufung herum daraufhin geprüft wurde, ob er genügend Material für mögliche Testfragen bietet, wurde als nächster Schritt das Sprach-Niveau der codierten Sinneinheiten bestimmt. Diese Einstufung erfolgte mithilfe des elektronischen Wörterbuchs „Profile deutsch“ (2002), das auf dem Gemeinsamen Europäischen Referenzrahmen des Europarates (2001) beruht und mit welchem rund 33’000 Einträge geprüft werden können. Mittels dieses Instrumentariums konnte bestimmt werden, dass die codierte Sinneinheit „differenziert“ (Steiner, 2001, S. 16, 2. Abschn., Zeile 2) auf dem Sprach-Niveau C liegt, „Differenzierung“ (Z. 5) aber liegt bereits auf Niveau D[24]. Ebenso auf Niveau D befindet sich die codierte Sinneinheit „(Verhaltens)-repertoire(s)“ (Z. 9). Beide Sinneinheiten wurden je für eine Aufgabenstellung verwendet (Aufgaben 1.1 und 1.2; siehe Abbildungen IV.5 und IV.3).

Wie oben erläutert, wurde bei der Konstruktion von Wortschatz-Items eine Fokussierung auf die codierte Sinneinheit priorisiert, während Aufgaben des allgemeinen oder detaillierten Leseverstehens sich am weiteren Kontext orientieren (mussten).

Die Aufgabe 1.2 (Abbildung IV.3), bei der es um die Kompetenz der Wortschatz-Beherrschung und -anwendung geht, fragt demzufolge wortgetreu die codierte Sinneinheit „-repertoire“ ab und prüft, ob das Fremdwort verstanden wird und durch ein Synonym ersetzt werden kann. Wie bei Wortschatz-Items häufig, erfolgt die Testung im geschlossenen Multiple-Choice-Format.

Abb. IV.3: Wortschatz-Aufgabe zum Originaltext in Abb. IV.4 und zu den codierten Sinneinheiten in Abb. IV.2

Aufgabe 1.2 **Welches Substantiv ersetzt in folgendem Kontext inhaltlich am besten das Fremdwort „...repertoire“?** **Vernachlässigen Sie die dadurch erforderlichen Beugungen.** „Wir können sagen, dass diese Reaktion zu den ursprünglichen und natürlichen Verhaltensmöglichkeiten gehört; sie ist Bestandteil eines zumindest in seinem Grundbestand angeborenen Verhaltensrepertoires.“ (Steiner [2001], S. 16) 1 ❑ ...mechanismus 2 ❑ ...vorrat 3 ❑ ...instinkt 4 ❑ ...weise	D_W_12

[24] Niveau D ist ein frei gewählter „Arbeitstitel“ für Wörter, welche im Grosswörterbuch Deutsch keinen Eintrag haben und demzufolge vermutlich einem höheren Anforderungsniveau als C1 und C2 entsprechen.

Abb. IV.4: Originaltext mit den codierten Sinneinheiten, den Niveau-Bestimmungen und der Textauswahl für die Aufgaben (Steiner, 2001, S. 16–17)

= D-Niveau ("Profile deutsch")
= C-Niveau

1.3 Klassisches Konditionieren von Emotionen

Die Situation

Eine junge Mutter sitzt mit ihrem knapp anderthalbjährigen Kind im Wartezimmer des Augenarztes. Nachdem letztes Mal die verstopften Tränenkanäle gespült werden mußten, sieht heute lediglich eine Nachkontrolle an. Das Kind ist viel unruhiger als sonst, aber die Mutter lenkt es mit Geschichtenerzählen geschickt ab. Wie nun die Arztgehilfin eintritt, beginnt das Kind wie am Messer zu schreien und zu strampeln. Das ist seine Reaktion auf die neue Reizsituation, die mit dem Eintreten der Arztgehilfin, die das Kind zuvor nicht kannte, eingetreten ist. Die Reaktion ist völlig eindeutig: sie ist der Ausdruck für eine *starke Emotion*, für *Angst*, für *Furcht* vor etwas, vielleicht auch für einen *Widerwillen* gegen etwas. Eine charakteristische Reizsituation löst eine ebenso charakteristische Reaktion aus. Verfolgen wir die Lerngeschichte dieser Reaktion!

Angstreaktionen als elementare Verhaltensweisen

Angst gehört zu den grundlegenden, schon früh zu beobachtenden Emotionen. Sie differenziert sich im Alter zwischen 3 und 6 Monaten zusammen mit dem *Zorn* und dem *Widerwillen* aus einer gemeinsamen Frühform, nämlich einem *Unbehagen* oder einer *Unlust* (engl. *distress*) heraus, die ihrerseits eine erste Differenzierung eines *allgemeinen Erregungszustandes* darstellt (nach Allport, 1961 bzw. Bridges, 1932). Wir können sagen, daß diese Reaktion zu den ursprünglichen und natürlichen Verhaltensmöglichkeiten gehört; sie ist Bestandteil eines zumindest in seinem Grundbestand angeborenen Verhaltenspertoires.

Daß es sich um eine höchst *elementare*, *natürliche Verhaltensweise* handelt, ist für die klassische Konditionierung von großer Bedeutung, denn diese erfolgt nur dort, wo bereits ausgebildete *elementare Verhaltensformen* (z.B. Reflexe wie der oben erwähnte Reflex der Speichelabsonderung) vorhanden sind, die durch neue Reize ausgelöst werden können.

Zunächst wenden wir uns der einfachen Frage zu: Welcher Reiz hat eigentlich das Weinen und Schreien des Kindes ausgelöst? – Wir müssen bekennen: Ganz genau können wir das gar nicht sagen! Wir wissen nur, daß sich mit dem Erscheinen der Arztgehilfin die *gesamte Reizsituation* verändert hat, und zwar in vielfältiger Weise. Die Person ist ein visuell wahrnehmbarer komplexer *Reiz* (engl. *stimulus*): eine menschliche Gestalt in weißem Kittel, mit einem Kopf von bestimmtem Aussehen, mit einem ganz bestimmten Körperbewegungs- und möglicherweise einem ganz bestimmten Annäherungsmuster. Sie ist aber auch, wenn sie spricht, eine Quelle für akustische Reize.

16

Auswahl von Stimuluselementen in komplexen Alltagssituationen

Die aktuelle Reizsituation in der Arztpraxis ist gegenüber derjenigen in Pawlows Labor eine viel komplexere! Durch welche *Teile* dieser Reizsituation nun eine Reaktion wie die Angstreaktion des Kindes ausgelöst wurde, ist in der Tat schwer zu sagen. Mit diesem Problem mußten sich die Lerntheoretiker der ersten Hälfte unseres Jahrhunderts schon auseinandersetzen. Sie hatten erkannt, daß in Alltagssituationen nur ausnahmsweise *einfache Einzelreize* auftraten, von denen man mit einiger Sicherheit annehmen konnte, daß sie eine bestimmte Reaktion auslösten. Der Lerntheoretiker Edwin R. Guthrie (1886–1959) war sich der Komplexität der Reizsituation im Alltag durchaus bewußt und hat von der Aktivität des Organismus bei der Stimulusauswahl gesprochen und formuliert, daß das, was der Organismus erfasse, zum Signal für das werde, was getan wird (Guthrie, 1959, 186). Das heißt nichts anderes, als daß nicht eine ganze Reizsituation eine Reaktion auslöst, sondern nur das, was aus ihr «erfaßt» wird; das aber wirkt dann als Signal und löst eine entsprechende Reaktion aus.

Schon 1931 hat ein anderer amerikanischer Lerntheoretiker, Edward L. Thorndike (1874–1949), Lerngesetze und -prinzipien formuliert, von denen eines das *unterschiedliche Hervortreten von Stimuluselementen* betraf. Eines der Nebenprinzipien befaßt sich mit der *Prävalenz von Elementen* (siehe Bower & Hilgard, 1981, 27). Dieses Prinzip besagt, daß ein Lerner selektiv auf besondere, d.h. herausragende oder auffallende Elemente der Reizsituation reagieren kann. Die auffallenden Teile der Reizsituation bieten sich dem Organismus gleichsam zum Erfassen an. Hier müssen wir einen bedeutenden Unterschied zwischen den Aussagen von Thorndike (1931) und Guthrie beachten: Während sich nach Thorndikes Ansicht die auffallenden Elemente einem passiven Organismus anbieten oder gar aufdrängen, spricht Guthrie ausdrücklich von der *Aktivität des Organismus* im Hinblick auf das, was von einer Reizsituation erfaßt werden soll. Guthrie weist also dem aktiven Organismus eine besondere Bedeutung gegenüber den Merkmalen der entsprechenden Reizsituation zu. Er sagt allerdings nichts Genaueres über diese Prozesse der Auslese von Reizmerkmalen; auch kommt er nicht darauf zu sprechen, daß allenfalls identische Teile von Reizsituationen für verschiedene Individuen völlig verschiedene Bedeutungen haben können und aus diesem Grund auch völlig unterschiedliche Reaktionen auslösen! Aufgrund der Pawlowschen Konditionierungstheorie oder der frühen Konnektionistischen Lerntheorien (wie denjenigen von Thorndike oder Guthrie) sind wir nicht in der Lage, präzise Vorhersagen über die Auswahl von Stimuluselementen (Teilen von Reizsituationen) zu machen; dazu sind Theorien notwendig, die etwas über die Aufmerksamkeitssteuerung zur Auswahl von Teilen von Reizen aussagen können, und das sind die *neueren Theorien aus der Kognitionspsychologie* (vgl. etwa Neisser 1976, 1979 oder Eysenck & Keane, 1990).

17

Die Aufgabe 1.1 (siehe Abbildung IV.5) fokussiert in ihrer Fragestellung ebenfalls die codierte Sinneinheit (in diesem Fall „differenzieren/Differenzierung"), bezieht sich aber auf den gesamten Textzusammenhang des Abschnittes. Die Antwortvarianten können nur erschlossen werden, wenn die Logik des Informationsstranges des Satzes verstanden und mit der konkreten Frage verknüpft wird. Hier bildet genaues Lesen des ganzen Kontexts die Voraussetzung für die richtige Antwort. Deshalb ist dieses Item im Kompetenzraster der Zelle des detaillierten Leseverständnisses zugeordnet. Hilfestellung für die Aufgabe bietet das beschriebene Faktum, dass Unlust eine emotionale Frühform darstellt. Der logische Fehler in Option 2 schliesst somit diese Antwort aus.

Abb. IV.5: Detailliertes Leseverstehen – Aufgabe zum Originaltext in Abb. IV.4 und zu den codierten Sinneinheiten in Abb. IV.2

Aufgabe 1.1 **In welchem Sinne ist *„differenzieren/Differenzierung"* in folgendem Textzusammenhang zu verstehen? Kreuzen Sie die richtige Antwort an.** „Angst gehört zu den grundlegenden, schon früh zu beobachtenden Emotionen. Sie differenziert sich im Alter zwischen 3 und 6 Monaten zusammen mit dem *Zorn* und dem *Widerwillen* aus einer gemeinsamen Frühform, nämlich einem *Unbehagen* oder einer *Unlust* (engl. distress) heraus, die ihrerseits eine erste Differenzierung eines *allgemeinen Erregungszustandes* darstellt (nach Allport, 1961 bzw. Bridges, 1932)." (Steiner, Gerhard [2001]. *Lernen: Zwanzig Szenarien aus dem Alltag.* Bern: Huber. Kap. 1: Angst vor weissen Kitteln – Klassisches Konditionieren, S. 16) 1 ❑ Angst entwickelt sich annäherungsweise zwischen drei und sechs Monaten. 2 ❑ Angst entwickelt sich simultan zu Zorn, Widerwille und Unlust. 3 ❑ Angst entwickelt sich primär als originäre Frühform. 4 ❑ Angst entwickelt sich bei der Aufgliederung eines allgemeinen Erregungszustandes.	D_D_11

Insgesamt wurden zehn unterschiedliche Test-Teile gebildet, die sich auf die verschiedenen in Teilprojekt A (siehe Kapitel 2) ausgewählten Studienfächer an Schweizerischen Universitäten beziehen. Jeder Testteil besteht aus mindestens einem Einstiegstext aus dem betreffenden Studienfach. Dieser Fachtext muss von den Probandinnen und Probanden in einem ersten Schritt gelesen werden; darauf aufbauend finden sich Testaufgaben in verschiedenen Formaten.

2.5.1.2 Besonderes bei der Übersetzung

Bei Erstsprache fällt für die Vergleichbarkeit der Testwerte besonders ins Gewicht, dass die den Aufgaben zugrunde liegenden Fachtexte durch die notwendige Übersetzung der Items in der einen Sprache grundsätzlich etwas einfacher oder schwieriger ausfallen können als in der anderen, und dass Bedeutungsnuancen, welche für ein bestimmtes Test-Item relevant sind, in der anderen Sprache weniger bedeutungsunterscheidend ausfallen können. Beispielsweise betrifft dies im Deutschen gewisse Fremdwörter, die an der Schwerverständlichkeit der Fachsprachen stark beteiligt sind und die im Unterschied dazu in der Romandie eher zur Alltagssprache gehören oder an vertrauteren Sprachwurzeln einfacher zu erkennen sind. Umgekehrt kann für die französisch- und italienischsprachigen Tests erschwerend dazukommen, dass die Primärtexte, die den Probandinnen und Probanden ausschliesslich als Übersetzungen vorlagen, nicht ihrem gewachsenen Sprachmodus entsprungen sind. Um eine genaue Aussage über die Einflussnahme solcher Faktoren auf die Testwerte geben zu können, wäre eine weiterführende textanalytische Untersuchung notwendig. In der Lese- und Schreibforschung liegen verschiedene, zum Teil ältere Methoden zur allgemeinen Schwierigkeitsbestimmung von Texten vor (Fucks, 1955; Björnsson, 1968; Groeben, 1972; Bamberger & Vanecek, 1984; Tauber, 1984; Drinkmann & Groeben, 1989). Heute erlebt bei-

spielsweise der in Skandinavien entwickelte Lesbarkeitsindex LIX von Björnsson (1968) dank computerbasierter Software für Textanalysen (z.B. das Programm TextQuest, Social Science Consulting, 2007) auch für den (Deutsch-)Unterricht und die Standarddiskussion eine Renaissance (Köster, 2005, S. 37), nachdem diese auf „durchschnittliche Satzlänge“ und „Prozentanteil Langwörter“ beruhende Verständlichkeitsformel bereits in den 1980er Jahren als „einfaches, objektives, aussagekräftiges“ Hilfsmittel zur Textbeurteilung propagiert wurde (Kommission für Schul- & Gemeindebibliotheken des Kantons Bern, 1989, S. 3). Inwiefern diese Analysen aber auch für universitäre Textsorten und insbesondere sprachvergleichend für Übersetzungen greifen, bleibt offen.

Das geplante Verfahren, in jeder Erstsprache spracheigene Items zu konstruieren und in die jeweils zwei anderen Landessprachen übersetzen zu lassen, konnte aus Ressourcengründen auch seitens der Projektpartner nicht eingehalten werden. Die erstmalige Testkonstruktion der Items erfolgte somit ausschliesslich im deutschsprachigen Team. Der Aufgabenpool Deutsch wurde anschliessend erstübersetzt und dann von den zweisprachigen Sprachexpertinnen der Projektpartner Romandie und Tessin geprüft und feinübersetzt. Gewisse Items mussten aus Sprachgründen von vornherein ausgeschieden beziehungsweise überarbeitet und angepasst werden; diese Endauswahl trafen die drei Zuständigen für Erstsprache gemeinsam.

Für die ausgewählten Items in Erstsprache wurden in allen drei Sprachregionen Pretests durchgeführt. Aufgrund der Resultate waren erste konkrete Sprachvergleiche und darauf basierend weitere Anpassungen möglich: Items oder Distraktoren, die im Tessin und in der Romandie wahrscheinlich anders verstanden wurden als im Original, wurden in einem gemeinsamen Diskussions- und Einigungsprozess weiter optimiert. Dieser Prozess der Abgleichung aller Sprach-Items für alle drei Erstsprachen fand sowohl, wie beschrieben, bei der Testkonstruktion als auch beim Korrekturschlüssel mehrstufig statt.

Eine weitere Besonderheit des Tests in Erstsprache bildet der so genannte Grammatik-Testteil, der auch Rechtschreibung und Interpunktion beinhaltet. Dieser Bereich kann aus sprachstrukturellen Gründen nicht mit übersetzten Items aus anderen Sprachen getestet werden, sondern muss per se ausschliesslich mit in der jeweiligen Landessprache entwickelten Test-Items erhoben werden. Aus Ressourcengründen verzichteten die Projektpartner aus der Romandie und dem Tessin darauf, für den Bereich Rechtschreibung und für weitere grammatikalische Kompetenzen eigene Test-Items zu konstruieren (die Tessiner Partner auch deshalb, weil sie grundsätzliche Bedenken gegen einen Grammatikteil hatten). Für die Erstsprache Deutsch wurde daraufhin entschieden, nicht zuletzt auch wegen der Befragungsergebnisse in Teilprojekt A4 wenigstens für die Erstsprache Deutsch nicht auf einen Testteil für Grammatik/Orthografie zu verzichten.

2.5.2 Beschreibung der Testhefte

Für Erstsprache Deutsch wurden aufgrund des gewählten Multi-Matrix-Designs fünf Testhefte mit rotierenden Aufgabenblöcken angelegt, für Französisch und Italienisch vier (ohne Grammatik). Somit sind in allen Sprachen drei Testhefte identisch; nur das vierte und fünfte sind aufgrund der zusätzlichen Grammatikaufgaben in Erstsprache Deutsch unterschiedlich.

Für den Haupttest wurden insgesamt 168 Items entwickelt. Die Blöcke der Testhefte für Erstsprache wurden aufgrund der folgenden inhaltlichen und testbedingten Kriterien zusammengestellt:

- In jedem Testheft sollen Einstiegstexte aus unterschiedlichen Studienfächern enthalten sein.
- Die Reihenfolge der Lese-Aufgaben ist unterschiedlich, das heisst, idealerweise ist ein Item gleich häufig am Anfang, in der Mitte und am Ende eines Heftes platziert. Auf diese Weise kann garantiert werden, dass Störvariablen wie Konzentrationsabfälle oder Leistungssteigerung durch „Einlesen“ nicht einseitig immer die gleichen Aufgaben treffen.
- Die Aufgaben sollen innerhalb von 22.5 Minuten zu bewältigen sein.

2.5.3 Charakterisierung der in die Hauptauswertungen einbezogenen Aufgaben und Verankerung im Kompetenzraster

Es wurden sowohl geschlossene als auch offene Formate entwickelt, um den unterschiedlichen Teilkompetenzen in Erstsprache gerecht zu werden. Von den insgesamt 124 in die Hauptauswertung übernommenen Test-Items sind 74 Items im offenen Format konstruiert und 50 Items im geschlossenen Multiple-Choice-Format (MC). Diese beiden Hauptformate „offen“ und „geschlossen“ werden weiter unterteilt in Unterformate oder Aufgabentypen. Von den 74 Test-Items im offenen Format sind 81% „offen kurz“: Hier handelt es sich um Aufgaben, bei denen die Maturandinnen und Maturanden z.B. ein Wort oder eine fehlende Endung in einen Lückentext einfügen mussten. Bei den 14 Test-Items im Unterformat „offen lang“ (18.6%) handelt es sich um anspruchsvolle Aufgabenstellungen, bei denen es z. B. darum ging, Zwischentitel für längere Abschnitte wissenschaftlicher Arbeiten zu setzen oder Sätze aus Sachtexten grammatikalisch richtig umzuformen. Von den Multiple-Choice-Items sind gut zwei Drittel im klassischen Format gehalten (Ankreuzen der richtigen Antwort). Daneben finden sich hier weitere Aufgabenformate, wie etwa Zuordnungsaufgaben (z. B. Zuordnung von vorgegebenen Titeln zu Textabschnitten).

In die Hauptauswertung wurden für Erstsprache insgesamt 77 „nationale“ Items einbezogen, die sowohl für Deutsch als auch für Französisch valide und reliabel waren. Für Erstsprache Deutsch wurden zusätzlich aus dem so genannten Grammatikteil 47 Teilaufgaben hinzugenommen, so dass der Auswertung für Deutsch insgesamt 124 Test-Items zur Verfügung standen. Sie verteilen sich wie folgt auf die Hauptkategorien des Kompetenzrasters:

- Allgemeines Leseverstehen (Zur Orientierung lesen): 19 Items (D_A_XX)
- Detailliertes Leseverstehen: Information und Argumentation verstehen: 40 Items (D_D_XX)
- Wortschatz: 19 Items (D_W_XX)
- Grammatische Kompetenz: 17 Items (D_G_XX)
- Orthografie: 29 Items (D_G_XX)

Die Testcharakteristiken sind im technischen Bericht beschrieben.

2.6 Der Test in Mathematik

2.6.1 Ergänzendes zur Testentstehung

2.6.1.1 Besonderes und Beispiel zur Aufgabenkonstruktion

Im Folgenden wird anhand eines Beispiels exemplarisch gezeigt, wie mithilfe der in den Studienunterlagen gefundenen Sinneinheiten die Testaufgaben konstruiert wurden.

In einem Skript ist der in der folgenden Abbildung IV.6 wiedergegebene Ausschnitt zu finden.

Abb. IV.6: Ausschnitt aus einem Skript (Quelle: ETH Zürich, Physik, Felder, Geometrie, S. 30)

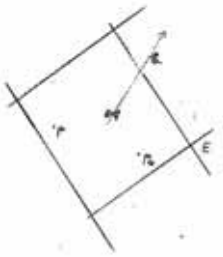
2.2 Raum 16.11.2004

Anwendung: Kartesische Gleichung einer Ebene. Ebene durch

$$P_0 = (x_0, y_0, z_0)$$

Normalvektor

$$\vec{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \neq \vec{0} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$



Ebene

$$= \{P \in \mathbb{R}^3 \mid \overrightarrow{P_0P} \perp \vec{a}\} = \left\{(x, y, z) \mid \begin{pmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = 0\right\}$$

$$= \{(x, y, z) \mid (x - x_0)a_1 + (y - y_0)a_2 + (z - z_0)a_3 = 0\}$$

Ebenengleichung:

$$xa_1 + ya_2 + za_3 = b$$

mit $b = x_0a_1 + y_0a_2 + z_0a_3$

Übung: Spiegelung s bezüglich der Ebene $\vec{a} \cdot \vec{x} = 0$ $\vec{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$

$$s(\vec{x}) = \vec{x} - \frac{2(\vec{a} \cdot \vec{x})}{\vec{a} \cdot \vec{a}} \cdot \vec{a}$$

2.2.2 Vektorprodukt

Definition: Vektorprodukt (Kreuzprodukt)

$$\mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$$

$$(\vec{u}, \vec{v}) \mapsto \vec{u} \times \vec{v}$$

$$\vec{u} \times \vec{v} = \begin{pmatrix} u_2v_3 - u_3v_2 \\ -u_1v_3 + u_3v_1 \\ u_1v_2 - u_2v_1 \end{pmatrix} = \begin{pmatrix} \det\begin{pmatrix} u_2 & v_2 \\ u_3 & v_3 \end{pmatrix} \\ -\det\begin{pmatrix} u_1 & v_1 \\ u_3 & v_3 \end{pmatrix} \\ \det\begin{pmatrix} u_1 & v_1 \\ u_2 & v_2 \end{pmatrix} \end{pmatrix}$$

Die wichtigen Sinneinheiten (vorausgesetztes Wissen, um den Skriptinhalt verstehen zu können) bestehen hier in der Art der Ebenenbeschreibung und der daraus abgeleiteten Koordinatengleichung („Ebenengleichung“). Um diese Ebenengleichung $xa_1 + ya_2 + za_3 = b$ zu verstehen, muss man wissen, dass der Vektor $\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$ senkrecht auf der definierten Ebene steht. Die entsprechende Testaufgabe findet sich in Abbildung IV.7

Abb. IV.7: Testaufgabe zum Lehrtext in Abb. IV.6

M_G_11

Für den Punkt $(x_0|y_0|z_0)$ und den Vektor $\vec{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$ stellt die Menge

$$\left\{ (x|y|z) \,\middle|\, \begin{pmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = 0 \right\}$$

eine Ebene im Raum dar, die den Punkt $(x_0|y_0|z_0)$ enthält.

Welcher Vektor steht senkrecht auf dieser Ebene?

1 $\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$

2 $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$

3 $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$

4 $\begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix}$

2.6.1.2 Besonderes bei der Übersetzung

Die Erstübersetzung in Italienisch erfolgte durch eine Übersetzerin und wurde durch den Autor der Aufgaben ein erstes Mal geprüft, bevor ein Mathematikdozent der Scuola Universitaria Professionale della Svizzera Italiana (SUPSI) sie erneut kontrollierte. Die Erstübersetzung in Französisch erfolgte durch den Autor der Aufgaben selbst. Die Überprüfung der Übersetzung leistete ein französischsprachiger und gleichzeitig bilingualer Mathematik-Fachdidaktiker für die Sekundarstufe II. Dabei erfolgten auch qualitative inhaltliche Anpassungen. Bei beiden Übersetzungen zeigte sich, dass es für gewisse mathematische Begriffe lokale Besonderheiten gibt, die nicht einmal in der gesamten Romandie beziehungsweise im Tessin identisch sind. Zudem unterscheiden sich einige Begriffe auch von jenen, die in Frankreich beziehungsweise Italien verwendet werden. In mehreren Fällen wurden zur Sicherheit weitere Experten angefragt oder einbezogen, deren Kommentare und Korrekturen schliesslich parallel eingearbeitet wurden; dies auch – sofern möglich und sinnvoll – bei der ursprünglichen deutschen Version. Die französischen Übersetzungen wurden darüber hinaus von zwei Schulklassen in Neuchâtel gelöst: Diese Testpersonen gaben keine Kommentare ab, die eine Korrektur der Testaufgaben nötig gemacht hätte.

2.6.2 Beschreibung der Testhefte

Für Mathematik wurden aufgrund des gewählten Multi-Matrix-Designs sechs Testhefte mit rotierenden Aufgabenblöcken angelegt.

Für den Haupttest lagen 93 Aufgaben mit insgesamt 175 Items vor. Bei der Verteilung der Aufgaben auf die Blöcke und damit auch auf die Testhefte wurden folgende Kriterien angewandt:

- Jeder Block sollte etwa gleich viele Aufgaben aus den vier Hauptgebieten Analysis, Geometrie, Stochastik und Vermischtes enthalten.
- Jeder Block sollte in etwa gleich viele Aufgaben der drei verwendeten Aufgabenformate (siehe nächster Abschnitt 2.6.3) enthalten.
- Der Zeitaufwand für die Bearbeitung sollte für alle Blöcke ungefähr gleich sein.

2.6.3 Charakterisierung der in die Hauptauswertungen einbezogenen Aufgaben und Verankerung im Kompetenzraster

Für den Haupttest wurden drei verschiedene Aufgabenformate verwendet:

(A) klassische Multiple-Choice-Aufgaben (MC-Aufgaben)

(B) Aufgaben mit mehreren Behauptungen

(C) offene Aufgaben mit eindeutigen Lösungen

Die in die Hauptauswertungen einbezogenen Aufgaben verteilen sich wie folgt auf das Kompetenzraster:

- 29 Items entfallen inhaltlich auf Analysis (A_XX), 35 Items auf Geometrie (G_XX), 24 Items auf Stochastik (S_XX) und 29 Items auf Vermischtes (X_XX).
- Bei den „Fähigkeiten“ ergibt sich folgende Verteilung:
  - Kenntnisse von Definitionen und Sätzen 25 Items
  - Qualitatives Verständnis mathematischer Begriffe 49 Items
  - Rechnen 27 Items
  - Operieren mit mathematischen Termen und Kalkülen 20 Items
  - Interpretieren von Diagrammen 20 Items
  - Textverständnis 13 Items
  - Bildliches Vorstellen 4 Items

Einige Items können verschiedenen Fähigkeitskategorien des Kompetenzrasters zugeordnet werden. Die Testcharakteristiken sind im technischen Bericht beschrieben.

2.7 Der Test in Biologie

2.7.1 Ergänzendes zur Testentstehung

2.7.1.1 Besonderes und Beispiel zur Aufgabenkonstruktion

Während die Relevanz mathematischer Kompetenzen für die allgemeine Studierfähigkeit allgemein unbestritten ist, erweist sich die Bedeutung der Verfügbarkeit von Wissen aus Biologie ausserhalb eines naturwissenschaftlichen Studiums als weniger evident. Deshalb wurden bei der Auswahl der Sinneinheiten, die Eingang in die Grundgesamtheit der Test-Items der Hauptuntersuchung für Biologie (126 Items) finden sollten, drei spezielle Kriterien der Repräsentativität angewandt: Im Sinne einer ersten Priorität wurden zuerst alle Sinneinheiten gesucht, welche besonders fächerübergreifend, also nicht nur für das Biologie- oder Medizinstudium von Bedeutung sind (Menge I). Dann kamen solche Sinneinheiten zum Zug, die an besonders vielen Textstellen gefunden wurden (Menge II), und schliesslich besonders anspruchsvolle Sinneinheiten, die eine Hürde beim Textverständnis und somit für die Studierfähigkeit darstellen dürften (Menge III). Falls aufgrund dieses Auswahlprozederes gewisse Sinneinheiten nicht Eingang in die Menge der Test-Items gefunden hatten, konnte sie der Fachdidaktiker trotzdem mit der Begründung einfliessen lassen, dass sie für die Biologie besonders zentrale Sinneinheiten darstellten („Klassiker“ des Begriffsgebäudes), deren Tiefe des Verständnisses im Rahmen der vorliegenden Studie unbedingt untersucht werden sollte. Dies war vor allem deshalb notwendig, weil die in Teilprojekt A ermittelten Sinneinheiten nicht alle Studienfächer aller Schweizer Universitäten vollständig abdecken (siehe Kapitel II).

Das folgende Beispiel zur Sinneinheit „Evolution“ stammt aus der Menge der fächerübergreifenden Quellen (siehe Abbildung IV.8; die handschriftlichen Markierungen zeigen die Codierarbeit). Es handelt sich um eine Seite aus einem Skript im Studienfach Publizistik und Kommunikationswissenschaft, welche die Sinneinheit „Evolution“ enthält.

Abb. IV.8: Seite aus einem Lehrtext im Studienfach Publizistik und Kommunikationswissenschaft

Ein weiterer, mit dem Emergenzbegriff aufs Engste verbundener Begriff, ist der des *sozialen Systems*. Auch dieser Begriff ist bereits zu einem frühen Zeitpunkt innerhalb der gesellschaftstheoretischen Tradition aufgetaucht, nämlich bei Herbert Spencer.
Ein als "System" bezeichneter Realitätsausschnitt (es kann sich hierbei bspw. um eine Organisation oder eine ganze Gesellschaft handeln) wird als Untersuchungsgegenstand bestimmt, um anschliessend dessen Elemente (soziale Handlungen, Kommunikation) samt ihren Eigenschaften und Beziehungen untereinander oder zum übergeordneten System zu untersuchen.

Soziales System nach Herbert Spencer

Exkurs "Herbert Spencer und das soziale System"
Herbert Spencer verwendet gleich wie Comte einen organischen Gesellschaftsbegriff. Darüber hinaus wird Gesellschaft bei Spencer, ebenfalls in Anlehnung an die Biologie, bereits explizit als "*emergentes System*" konzipiert. Das soziale System "Gesellschaft" gleicht dem biologischen System dadurch, dass es aus lebenden Teilen besteht. Die Teile, gemeint sind Menschen in ihrer Zugehörigkeit zu sozialen Schichten, Berufen, Alterklassen usw., sind einander ungleich und übernehmen unterschiedliche Funktionen. Durch die Erfüllung ihrer unterschiedlichen Funktionen ermöglichen die Teile das Fortbestehen des Ganzen und der übrigen Teile. Die Qualität und die Dauerhaftigkeit der arbeitsteiligen Beziehungen zwischen den Teilen macht die Stabilität der emergenten Ordnung des Ganzen aus. Stabilität ist bei Spencer ein zentrales Kennzeichen sowohl organischer wie auch sozialer Systeme.
Bei Spencer kommt noch ein weiteres Bestimmungsmoment sozialer Systeme hinzu - ein Moment, das bei Comte in dieser Weise noch nicht zu finden ist: gleich wie organische Systeme unterstehen soziale Systeme den Prinzipien der *Evolution*. Der Prozess der sozialen Evolution entspricht Spencer zufolge einem kontinuierlichen gesellschaftlichen Wachstumsprozess. Die evolutionäre Entwicklung in Richtung grösserer Komplexität untersteht dabei den gleichen Selektionskriterien wie die biologische Evolution nach Darwin. Herbert Spencer gilt denn auch neben John Stuart Mill als Begründer des ebenso bekannten wie problematischen Sozialdarwinismus (vgl. Spencer 1914). Der populäre Leitsatz "Survival of the Fittest" stammt im übrigen nicht von Charles Darwin, sondern von Herbert Spencer.

3!

Die Sinneinheit „Evolution“ tritt viermal auf.

5

Bei der Itemkonstruktion musste sich der Testautor zunächst die Frage stellen, in welcher Tiefe hier die Leserin/der Leser das Konzept der Evolution beziehungsweise der Evolutionstheorie verstanden haben muss, um die Passage ohne Rückgriff auf ein Wörterbuch zu verstehen. Die Antwort lautet: Sicherlich ist kein derart tiefes Verständnis nötig, dass die nachfolgend angedeuteten zwei grossen Theoriepfeiler zum Begriff Evolution aus dem Biologieunterricht erinnert werden müssen, um dem Text folgen zu können (nach Mayr, 2001):

a) Eine Population oder Art verändert sich dadurch, dass sie beständig neue Genotypen hervorbringt, und dadurch, dass die meisten Genotypen einer Generation wieder eliminiert werden.

b) Elimination: Entweder überlebt der Phänotyp eines Genotyps nicht (natürliche Selektion) oder der Phänotyp eines Genotyps hat nur geringen oder gar keinen Fortpflanzungserfolg (sexuelle Selektion).

Im genannten Textbeispiel dürften folgende (implizite) Assoziationen vollkommen ausreichen:

1. Der Prozess der Evolution arbeitet „blind“, das heisst, er läuft ab, ohne ein Ziel zu erreichen oder erreichen zu „wollen“ (vgl. Option 1 der konstruierten Aufgabe) beziehungsweise ohne von einer äusseren Macht in Richtung Ziel „gesteuert“ zu werden (vgl. Option 3).

2. Die Beschreibung des Phänomens der Evolution ist noch nicht dasselbe wie eine Theorie, welche die dahinterstehenden Prozesse so gut wie möglich zu beschreiben versucht (vgl. Option 4).

3. Ohne Variation gibt es keine Selektion (vgl. Option 2).

Daraus dürfte die Leserin beziehungsweise der Leser (implizit) schliessen, dass die soziale Evolution

1. nicht auf eine bestimmte (ideale) Gesellschaftsform hinauslaufen muss,
2. nicht unbedingt schon in einer Theorie der gesellschaftlichen Weiterentwicklung erklärt sein muss, wenngleich das Phänomen der gesellschaftlichen Weiterentwicklung weltweit gewissermassen „täglich“ zu beobachten ist,
3. auf einer Vielfalt verschiedener Ideen des Zusammenlebens aufbaut (Variation), die sich verbreiten (positive Selektion), sich über eine bestimmte Zeit erhalten und mit der Zeit wieder verschwinden kann (negative Selektion im Sinne einer Elimination).

Falls alle drei Assoziationen gemacht werden, hat der Autor des Lehrtextes an dieser Stelle mit dem Rückgriff auf den biologischen Begriff beziehungsweise die Metapher der Evolution Bedeutendes über seine Vorstellung der gesellschaftlichen Entwicklung aussagen können, ohne viele Worte verlieren zu müssen.

Die konkrete Ausformulierung des Items überprüft nun, ob die Probandin beziehungsweise der Proband beim Lesen des Skriptes keine „unglücklichen“ Assoziationen machen würde, die vom Autor der Lehrunterlage nicht beabsichtigt sind (siehe Abbildung IV.9):

Abb. IV.9: Testaufgabe zum Skriptausschnitt in Abb. IV.8

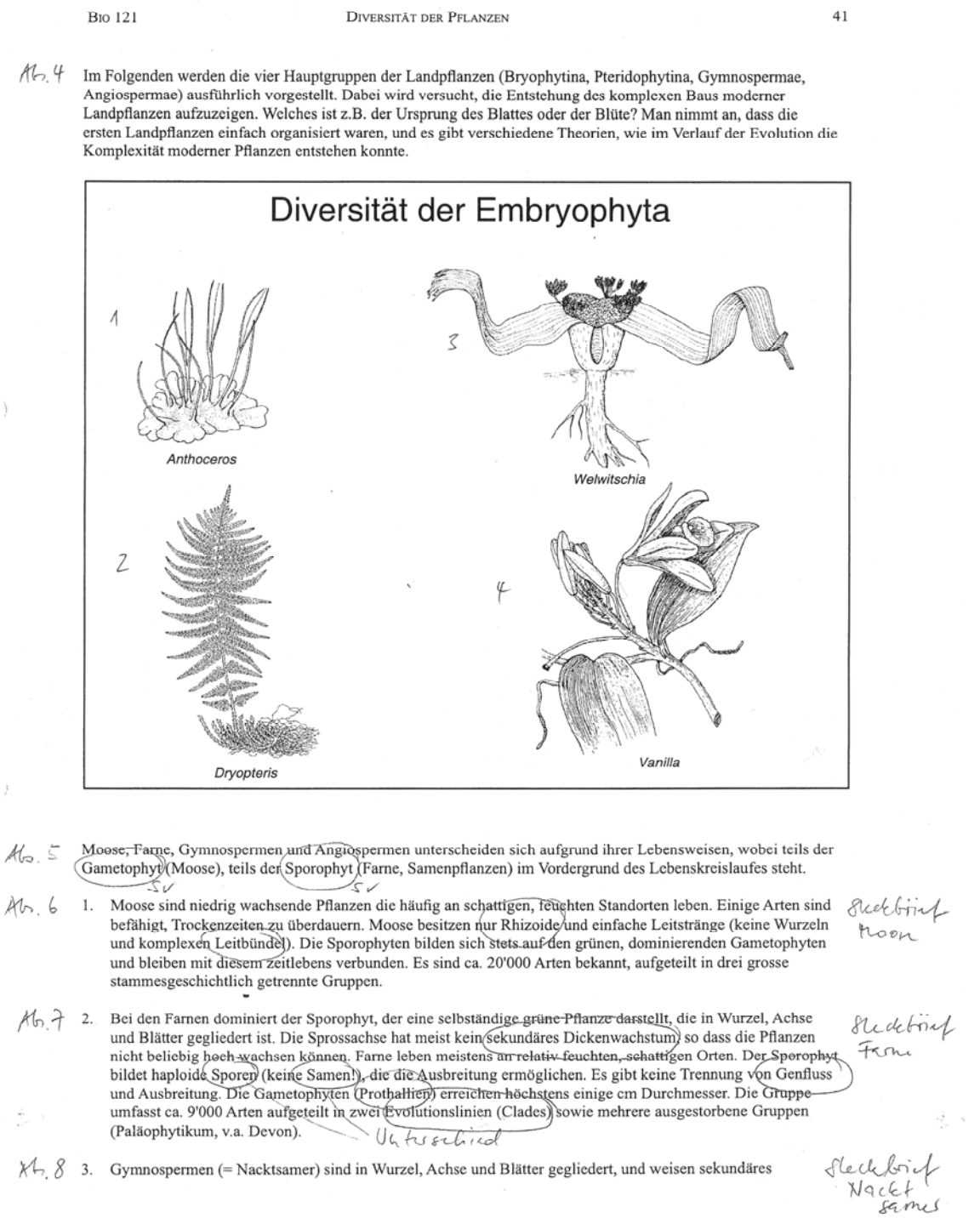
Test-Item B_EVO_01

Welches ist die zentrale Vorbedingung für jede Art von Evolution?

- ❑ Es muss ein Ziel geben, wohin die Evolution laufen soll.
- ❑ Es muss eine Vielfalt geben und eine Auswahl möglich sein.
- ❑ Es muss eine äussere Macht geben, welche die Vorgänge der Evolution steuert.
- ❑ Es muss eine Theorie geben, wie die Vorgänge der Evolution ablaufen müssen.

Ein weiteres Beispiel bezieht sich auf die Sinneinheiten „Sporen“ und „Samen“ aus der Menge „anspruchsvoller Quellen“ aus dem Biologiestudium. Der Lehrtext baut darauf auf, dass eine wesentliche Begriffsunterscheidung bereits im voruniversitären Biologieunterricht vorgenommen und im Langzeitgedächtnis abgelegt wurde. Im Skript „Biodiversität: Pflanzen, Wirbeltiere, Pilze (Teil Pflanzen)“ kann die Notwendigkeit des Klammerausdrucks in Absatz 7 auf Seite 41 (siehe Abbildung IV.10) nur erfasst werden, wenn der Unterschied zwischen Samen und Sporen bekannt ist.

Abb. IV.10: Seite aus einem Lehrtext im Studienfach Biologie



BIO 121 DIVERSITÄT DER PFLANZEN 41

Abb. 4 Im Folgenden werden die vier Hauptgruppen der Landpflanzen (Bryophytina, Pteridophytina, Gymnospermae, Angiospermae) ausführlich vorgestellt. Dabei wird versucht, die Entstehung des komplexen Baus moderner Landpflanzen aufzuzeigen. Welches ist z.B. der Ursprung des Blattes oder der Blüte? Man nimmt an, dass die ersten Landpflanzen einfach organisiert waren, und es gibt verschiedene Theorien, wie im Verlauf der Evolution die Komplexität moderner Pflanzen entstehen konnte.

Abb. 5 Moose, Farne, Gymnospermen und Angiospermen unterscheiden sich aufgrund ihrer Lebensweisen, wobei teils der Gametophyt (Moose), teils der Sporophyt (Farne, Samenpflanzen) im Vordergrund des Lebenskreislaufes steht.

Abb. 6 1. Moose sind niedrig wachsende Pflanzen die häufig an schattigen, feuchten Standorten leben. Einige Arten sind befähigt, Trockenzeiten zu überdauern. Moose besitzen nur Rhizoide und einfache Leitstränge (keine Wurzeln und komplexen Leitbündel). Die Sporophyten bilden sich stets auf den grünen, dominierenden Gametophyten und bleiben mit diesem zeitlebens verbunden. Es sind ca. 20'000 Arten bekannt, aufgeteilt in drei grosse stammesgeschichtlich getrennte Gruppen. Steckbrief Moos

Abb. 7 2. Bei den Farnen dominiert der Sporophyt, der eine selbständige grüne Pflanze darstellt, die in Wurzel, Achse und Blätter gegliedert ist. Die Sprossachse hat meist kein sekundäres Dickenwachstum, so dass die Pflanzen nicht beliebig hoch wachsen können. Farne leben meistens an relativ feuchten, schattigen Orten. Der Sporophyt bildet haploide Sporen (keine Samen!), die die Ausbreitung ermöglichen. Es gibt keine Trennung von Genfluss und Ausbreitung. Die Gametophyten (Prothallien) erreichen höchstens einige cm Durchmesser. Die Gruppe umfasst ca. 9'000 Arten aufgeteilt in zwei Evolutionslinien (Clades) sowie mehrere ausgestorbene Gruppen (Paläophytikum, v.a. Devon). Steckbrief Farne

Abb. 8 3. Gymnospermen (= Nacktsamer) sind in Wurzel, Achse und Blätter gegliedert, und weisen sekundäres Steckbrief Nacktsamer

Erstsemestrige der in Teilprojekt A1 untersuchten Studienrichtung Biologie müssen dieses Wissen mitbringen, wenn sie hier nicht eine wichtige Stelle überlesen wollen. Der Text hebt mit Ausrufezeichen einen wichtigen Unterschied hervor, erklärt ihn aber nicht. Die Kenntnis des Unterschieds zwischen Samen und Sporen stellt somit eindeutig Eingangswissen dar.

Das Item in Abbildung IV.11 testet, ob die Maturandin beziehungsweise der Maturand als Studienanfängerin beziehungsweise Studienanfänger in der Lage wäre, das Augenmerk auf das bei der Unterscheidung zwischen Samen und Sporen relevante Merkmal zu richten (nämlich die Zahl der Zellen, was zum Ausschluss der Optionen 3 und 4 führt) und ob er/sie Sporen als im Zweifelsfall einfacher organisierte (nämlich einzellige) Fortpflanzungseinheit einstufen würde (führt zum Ausschluss der Option 1). Eine Hilfestellung bietet auch der fachliche Fehler in Option 1: Organe sind immer mehrzellig, da sie aus Geweben bestehen und Gewebe per definitionem mehrzellig sind.

Abb. IV.11: Testaufgabe zum Lehrtext in Abb. IV.10

Test-Item B_GEN_16

Wie unterscheiden sich Sporen von Samen?

- ❑ Sporen sind immer mehrzellig; Samen sind einzellige Organe der Samenpflanzen.
- ❑ Sporen sind immer einzellig; Samen sind mehrzellige Organe der Samenpflanzen.
- ❑ Sporen werden von Moosen, Farnen und Bäumen produziert; Samen werden von Samenpflanzen produziert.
- ❑ Samen werden von allen Landpflanzen produziert; Sporen werden nur von Algen produziert.

2.7.1.2 Besonderes bei der Übersetzung

Hier wurden nach der Übersetzung durch je zwei Übersetzungspersonen die beiden Versionen in jeder Sprache von einer weiteren Person (ebenfalls mit Muttersprache in der Zielsprache und mit Fachstudium Biologie) verglichen und zu einer Schlussversion verarbeitet, und zwar durch Zusammenfügen der jeweils besten Textblöcke aus beiden Versionen.

In einem letzten Schritt wurde die Qualität der französischen Übersetzungen direkt an Maturandinnen und Maturanden mittels eines qualitativen Vorgehens erprobt. Zwei Schülerinnen und ein Schüler aus dem Collège Du Sud in Bulle bearbeiteten sämtliche Items. Dabei anwesend waren der Autor der deutschsprachigen Items sowie einer der Erstübersetzer. Diese beiden Personen sammelten zu jedem Item fortlaufend Rückmeldungen: einerseits zur sprachlichen Qualität der Fragestellungen, andererseits aber auch dahingehend, ob die richtige Option innerhalb der Multiple-Choice-Fragestellungen für Probanden mit entsprechendem Fachwissen klar erkennbar sei. Dabei stellte sich heraus, dass nur noch sehr vereinzelt sprachliche Feinheiten verbessert werden konnten, hinsichtlich des wesentlich wichtigeren Kriteriums der eindeutigen Erkennbarkeit der richtigen Option hingegen schon das Optimum erreicht war. Aus Zeitgründen (und wegen methodischer Einwände der Tessiner Partner) sowie aufgrund der Annahme, dass auch in der italienischsprachigen Übersetzung zumindest die eindeutige Erkennbarkeit der richtigen Optionen gegeben sei, wurde von einer analogen qualitativen Validierung der Items mit italienisch sprechenden Maturandinnen und Maturanden abgesehen.

2.7.2 Beschreibung der Testhefte

Für Biologie wurden aufgrund des gewählten Multi-Matrix-Designs sieben Testhefte mit rotierenden Aufgabenblöcken angelegt (vgl. Tabelle IV.2 in Abschnitt 2.1).

Bei der Verteilung der Aufgaben auf die Blöcke und damit auch auf die Testhefte wurden folgende Kriterien angewandt:

- Jeder Block soll je drei Aufgaben aus den sieben Themengebieten enthalten.
- Der Zeitaufwand sollte für alle Blöcke ungefähr gleich sein.

2.7.3 Charakterisierung der in die Hauptauswertungen einbezogenen Aufgaben und Verankerung im Kompetenzraster

Von den 91 Items der gelockerten Auswahlvariante sind 85 im Multiple-Choice-Format mit vier Antwortoptionen und exakt einer fachlich korrekten Option formuliert. Sechs Items sind offen formuliert und konnten mit einem biologischen Fachbegriff beantwortet werden.

Die ausgewählten Items verteilen sich wie folgt auf die Kategorien des Kompetenzrasters:

- Cytologie/Anatomie/Stoffwechsel: 31 Items (B_FKT_XX, B_STW_XX)
- Informationsverarbeitung, Verhalten und Immunbiologie: 17 Items (B_INF_XX)
- Genetik/Entwicklungsbiologie: 18 Items (B_GEN_XX)
- Ökologie: 15 Items (B_OEK_XX)
- Systematik/Evolution: 10 Items (B_EVO_XX).

Bei den Anforderungsbereichen ergibt sich folgende Verteilung:

- Stufe I: 47 Items
- Stufe II: 44 Items

Items auf Stufe III fehlen deshalb, weil deren Konstruktion nur im ausführlichen Essayformat möglich gewesen wären (Fähigkeiten im Bereich des eigenständigen, kreativ-schöpferischen, mehrstufig-iterativen Bearbeitens von komplexen Problemen; beispielsweise ein experimentelles Design zur Beantwortung einer spezifischen Fragestellung entwerfen, oder über einen biologischen Sachverhalt kritisch nachdenken). Die Bewertung und Codierung von Essays mit hohem kreativem Gehalt wäre sehr ressourcenaufwändig und wenig objektivierbar gewesen.

Im Gegensatz zu den Items in Erstsprache und Mathematik verteilen sich die zugrunde liegenden Studientexte weniger gut auf alle in Teilprojekt A1 und A3 untersuchten Studienfächer, weil Biologie im Vergleich zu diesen Unterrichtsfächern am wenigsten auf viele oder gar alle Studienfächer (mit)vorbereitet. 81 von 91 Test-Items betreffen biologisches Wissen, das nur für die analysierten Studienfächer Biologie und Medizin als Eingangswissen codiert wurde. 38 Items repräsentieren dabei das Eingangwissen, das für ein Studium speziell der Biologie oder Medizin *besonders häufig* vorausgesetzt wird, und 43 Items basieren auf besonders anspruchsvollen Sinneinheiten beziehungsweise vorausgesetztem Wissen aus den Studienmaterialien dieser Fächer.

Die Testcharakteristiken sind im technischen Bericht beschrieben.

2.8 Der überfachliche Fähigkeitstest (UFT)

2.8.1 Beschreibung der Grundlagen

Der in der Schweiz als Zulassungsprüfung zum Medizinstudium eingesetzte Medizin-Eignungstest (EMS) besteht aus den folgenden Untertests:

1. Quantitative und formale Probleme
2. Konzentriertes und sorgfältiges Arbeiten
3. Medizinisch-naturwissenschaftliches Grundverständnis
4. Muster zuordnen
5. Diagramme und Tabellen
6. Textverständnis
7. Fakten lernen
8. Schlauchfiguren
9. Figuren lernen

Am prognose-relevantesten für den Studienerfolg sind die vier Untertests 1, 3, 5 und 6 (Hänsgen & Spicher, 2001). Aus diesen wählte das für den Medizin-Eignungstest (EMS) in der Schweiz zuständige Zentrum für Testentwicklung und Diagnostik (ZTD) hinsichtlich Schwierigkeit und Trennschärfe angemessene Items aus und stellte vier Blöcke für eine Bearbeitungszeit von je 22.5 Minuten zusammen.

2.8.2 Beschreibung der Testhefte

Es wurden aufgrund des gewählten Multi-Matrix-Designs vier Testhefte mit rotierenden Aufgabenblöcken angelegt. Jedes Testheft bestand aus zwei Blöcken, ausgewählt aus den folgenden:

Block 1: Quantitative und formale Probleme (U_Q), 10 Aufgaben (UFT 1)

Block 2: Textverständnis (U_T), 12 Aufgaben (UFT 2)

Block 3: Naturwissenschaftliches Grundverständnis (U_N), 10 Aufgaben (UFT 3)

Block 4: Diagramme und Tabellen (U_D), 10 Aufgaben (UFT 4).

2.8.3 Charakterisierung der in die Hauptauswertungen einbezogenen Aufgaben

Wie bei den Fachtests wurden auch die Aufgaben des überfachlichen Fähigkeitstests auf die Erfüllung der in Abschnitt 2.4 beschriebenen teststatistischen Mindestanforderungen analysiert.

Alle Aufgaben erfüllen diese Kriterien. Es zeigte sich, dass es sich um bereits in anderem Zusammenhang gut erprobte Items handelt. Weil der UFT nur in der Deutschschweiz eingesetzt wurde, erübrigte sich eine Analyse möglicher Übersetzungsdifferenzen.

Es handelt sich bei allen Aufgaben um geschlossene Aufgabenformate. Die Testcharakteristiken sind im technischen Bericht beschrieben.

2.9 Der Zusatzfragebogen

Der Zusatzfragebogen wurde von den Maturandinnen und Maturanden im Anschluss an die Bearbeitung der Tests ausgefüllt; sein Einsatz bildete insofern den Abschluss der gesamten Erhebung im Rahmen des Teilprojekts C. Da in den deutschsprachigen Schulen mehr Zeit zur Verfügung stand, konnten in die deutsche Version einige zusätzliche Fragen aufgenommen werden. Der Fragebogen wurde in deutscher Sprache konzipiert, und anschliessend wurden die in allen Regionen erhobenen Items ins Französische und Italienische übersetzt. Insofern ist insbesondere beim Einsatz von Skalen jeweils separat für jede Region die Reliabilität zu prüfen (siehe dazu den technischen Bericht).

2.9.1 Begründung der Zusatzerhebung und Inhalte

Der Zusatzfragebogen wurde zum einen eingesetzt, um erweiterte Informationen zur Stichprobe zu erhalten, die über die von den Schulen bereitgestellten administrativen Daten hinausgehen. Dazu zählen vor allem Daten zur sozialen Herkunft der Familien, zu einem potenziellen Migrationshintergrund und den Sprachgewohnheiten der Maturandinnen und Maturanden. Diese Indikatoren stellen wichtige Kontrollvariablen in Bezug auf Schul- beziehungsweise Testleistungen dar. Zum anderen wurden einige Fragen zu ausgewählten überfachlichen Kompetenzen sowie Aspekten der Persönlichkeit und der Motivation (siehe Abschmitt 1.2.1) in den Fragebogen aufgenommen, die für die Studierfähigkeit besonders relevant sind. Aufgrund der Ressourcenbeschränkung konnten diese nur per schriftliche Befragung und damit lediglich über den Weg der Selbstauskunft der Maturandinnen und Maturanden erhoben werden, also nicht über objektivere Verfahren wie Assessment oder Beobachtung. Ausgewählt wurden dabei insbesondere Kompetenzen, die nicht bereits Gegenstand von EVAMAR I (Ramseier et al., 2004) oder der Studie zum Übergang vom Gymnasium ins Studium (Notter & Arnold, 2006) sind, und für die

ein In-Beziehung-Setzen zu den Testleistungen lohnenswert erscheint. Des Weiteren wurden die Maturandinnen und Maturanden nach ihrer Wahrnehmung der Tests und zu ihren Plänen nach der Maturität befragt. Für die deutschsprachigen Maturandinnen und Maturanden schliesslich wurden zusätzlich das Interesse am Unterricht und das Ausmass des häuslichen Arbeitsaufwands (jeweils bezüglich der getesteten Schulfächer Mathematik, Biologie und Deutsch), einige Fragen zur studienpropädeutischen Funktion des Gymnasiums und zum Leseinteresse in den Fragebogen aufgenommen. Im Folgenden werden die einzelnen Bereiche genauer beschrieben.

2.9.2 Beschreibung des Fragebogens

Um zu prüfen, wie ernsthaft die Maturandinnen und Maturanden die Tests bearbeitet hatten, enthielt der Fragebogen zwei Dimensionen: Zum einen wurde erfragt, wie interessant sie die verschiedenen Tests fanden (vierstufige Skala von 1 = nicht interessant bis 4 = sehr interessant), zum anderen diente eine vierstufige Skala der Erhebung der Anstrengungsbereitschaft der Maturandinnen und Maturanden, die Tests sorgfältig, engagiert und konzentriert zu bearbeiten.

Im Bereich „Schule und Unterricht“ ist neben der Frage nach der Belegung von Schwerpunkt- und Ergänzungsfach eine Reihe von Items zu verschiedenen überfachlichen Kompetenzen angesiedelt. Die Maturandinnen und Maturanden sollten jeweils beurteilen, inwiefern Aussagen auf sie zutreffen, die folgende Kompetenzen und Lernstrategien thematisieren: „Zeitplanung“, „Prüfungsvorbereitung“, „Zeiteffizienz“, „Konzentrationsfähigkeit“, „Ausdauer“, „Vernetztes Denken“, „Diskussionsfähigkeit“ und „Perspektivenübernahme“.

In der Version für die deutschsprachigen Maturandinnen und Maturanden finden sich im Bereich „Schule und Unterricht“ darüber hinaus Fragen bezüglich des Zeitaufwands für das häusliche Lernen für die Fächer Mathematik, Deutsch und Biologie sowie Items, die das Interesse und die Motivation bezüglich der drei Testfächer erfassen sollten. Dabei sollten die Maturandinnen und Maturanden beurteilen, in welchem Ausmass sie Aussagen wie z. B. „im Mathematik-Unterricht macht das Lernen Spass“ zustimmen können (von 1 = trifft nicht zu bis 4 = trifft völlig zu). Weiterhin sollten die Maturandinnen und Maturanden angeben, inwieweit sie in der Schule verschiedene studienpropädeutische Tätigkeiten ausführen (wie z.B. „Informationen für eine Arbeit finden, auswählen und einordnen“) beziehungsweise wie häufig sie in den letzten zwei Jahren Tätigkeiten wie z.B. „Wissenschaftliche Experimente planen“ ausgeübt haben. Ein weiterer Komplex von Fragen bezieht sich auf die wahrgenommene Vorbereitung auf Anforderungen des Studiums. Hier sollten die Gymnasiastinnen und Gymnasiasten angeben, inwiefern sie Aussagen wie beispielsweise „Ich glaube, ich habe im Gymnasium verstanden, wie man Fragestellungen wissenschaftlich bearbeitet“ oder „Ich habe am Gymnasium gelernt, Texte systematisch zu analysieren“ zustimmen können.

Der nächste Abschnitt von Fragen bezieht sich auf die Pläne der Maturandinnen und Maturanden nach der Maturität. Alle Probandinnen und Probanden wurden nach der Absicht gefragt, ein Studium aufzunehmen und – bei Vorhandensein eines solches Planes – nach dem angestrebten Studienfach und der dafür gewählten beziehungsweise gewünschten Institution. Auch wurde nach etwaigen Plänen für ein Auslandsstudienjahr gefragt. Maturandinnen und Maturanden, die hingegen kein Studium aufnehmen wollen, sollten ebenfalls ihre beruflichen Zukunftspläne offenbaren. Im Fragebogen für die deutschsprachigen Maturandinnen und Maturanden wird in diesem Rahmen schliesslich nach der Teilnahme am Eignungstest für das Medizinstudium (EMS) und eventuellen Trainingsmassnahmen im Hinblick auf eine geplante Teilnahme gefragt.

Den Abschluss des Fragebogens bilden Fragen zur Person: Neben dem Geschlecht sollten die Gymnasiastinnen und Gymnasiasten angeben, wo sie und ihre Eltern geboren sind, wie lange die im Ausland geborenen in der Schweiz leben und welche Sprache sie überwiegend im Elternhaus und im Freundeskreis sprechen. Fragen zum letzten Bildungsabschluss der Eltern, zum Buchbestand im familiären Haushalt sowie zum Vorhandensein kultureller Wohlstandsgüter wie beispielsweise einem „Zimmer für sich allein“ oder dem Vorhandensein von Kunstwerken dienen der Beschreibung der sozialen Herkunft der Versuchspersonen. Ausschliesslich in der Version für die deutschsprachige Schweiz schliessen sich zusätzliche Fragen zu Arbeiten neben der Schule und zum Leseinteresse der Maturandinnen und Maturanden an.

2.9.3 Güte des Instruments

Zum Teil wurde auf bereits bestehende, bei anderen Untersuchungen wie PISA (vgl. Kunter et al., 2002), BIJU (Bildungsverläufe und psychosoziale Entwicklung im Jugend- und jungen Erwachsenenalter) oder TOSCA (vgl. jeweils Trautwein et al., 2006) bereits eingeführte Instrumente und Skalen zurückgegriffen, zum Teil wurden neu konstruierte Items im Rahmen der Pretests (vgl. Abschnitt 2.2) erprobt und für den Haupttest entsprechend modifiziert. Bis auf wenige Ausnahmen weisen die eingesetzten Skalen zufrieden stellende interne Konsistenzen (Cronbachs Alpha) auf, die sich meist in ähnlicher Grössenordnung bewegen wie bei den Originalskalen. Erfreulicherweise zeigt sich dieses Ergebnis auch bei jeweils separater Prüfung der verschiedenen Regionen: Die internen Konsistenzmasse ähneln sich über die Teilstichproben und die verschiedenen Skalen hinweg stark, obwohl die Items, wie bereits beschrieben, für den Einsatz im Rahmen von EVAMAR II erstmals übersetzt wurden.

3 Bestimmung der Stichprobe

3.1 Bestimmung der Grundgesamtheit

Zur Grundpopulation gehören die Gymnasiastinnen und Gymnasiasten der Schweiz, die im Sommer 2007 die Maturitätsprüfungen abgelegt haben. Von der Population ausgeschlossen waren die Gymnasiastinnen und Gymnasiasten der Kantone Basel-Landschaft und Genf. Im Kanton Basel-Landschaft findet die Maturitätsprüfung jeweils erst im Herbst statt, weshalb die Vergleichbarkeit nicht gegeben war. Die Gymnasien des Kantons Genf erhoben schon früh massive Einwände gegen die Untersuchung, und die meisten Schulen verweigerten die Übermittlung der Angaben, die Ende 2006 eingeholt und zur Bestimmung der Grundgesamtheit notwendig gewesen wären. Die Steuergruppe des Projekts EVAMAR II verzichtete in der Folge aus verschiedenen Gründen darauf, eine allfällige, von den zuständigen Behörden von oben verordnete Teilnahme zu erwirken. Eine zwangsweise Teilnahme hätte ein erhebliches Verfälschungspotenzial für die Validität der Ergebnisse mit sich gebracht.

Um einerseits die heterogene Struktur der Schweizer Gymnasiallandschaft möglichst gut abzubilden und andererseits verschiedene Vergleiche von Subpopulationen zuverlässig durchführen zu können, wurde die Stichprobe in insgesamt sieben explizite Einheiten stratifiziert. Die Straten-Einteilung erfolgte entsprechend folgender Überlegungen: Zürich als grösster Kanton stellt eine Einheit dar, für die sich aufgrund der Grösse auch interne strukturelle Auswertungen vornehmen lassen, ohne dass eine Vollerhebung notwendig ist[25]; als eine weitere Gruppe bot sich innerhalb der Deutschschweiz der deutschsprachige Teil des Kantons Bern an, in dem dort bis zum Maturitätsabschluss – als strukturelle Eigenheit – das erste Jahr von vier Jahren an einer Sekundarschule stattfinden kann und nur drei Jahre im eigentlichen Gymnasium absolviert werden müssen; „Deutschschweiz klein“ umfasst die kleinen Deutschschweizer Kantonen, die weniger als 15 Maturaklassen haben (AI, NW, OW, GL, UR, AR, SH, VSd, ZG) – damit sollte deren sonst kleinere Auswahlwahrscheinlichkeit berücksichtigt werden; „Deutschschweiz gross“ besteht aus den restlichen Deutschschweizer Kantonen, die mehr als 15 Klassen haben (SZ, SO, TG, GR, BS, AG, SG, LU); eine Einheit bildet „Romandie 1“ mit den Westschweizer Kantonen, in denen ebenfalls nur drei Jahre am eigentlichen Gymnasium absolviert werden müssen (BEf, JU, NE, VD), und „Romandie 2“ mit den Westschweizer Kantonen mit mindestens vier Jahren Ausbildungsdauer am eigentlichen Gymnasium (FR, VSf); als eigenes Stratum war schliesslich auch das Tessin vorgesehen.

Tabelle IV.3 zeigt die Anzahl Klassen sowie Maturandinnen und Maturanden der Grundgesamtheit, aufgeteilt nach Stratenbildung.

[25] Diese Auswertungen werden ergänzend vorgenommen und sind nicht durch den Auftraggeber EDK/SBF finanziert.

Tab. IV.3: Grundgesamtheit

Stratum	Definition	Anzahl Klassen	Anzahl Maturandinnen und Maturanden
1	Deutschschweiz 1: Zürich	114	2144
2	Deutschschweiz 2: Bern deutsch (drei Mindest-Jahre am Gymnasium)	98	1927
3	Deutschschweiz 3: kleine Kantone (Klassenzahl <= 15): AI, NW, OW, GL, UR, AR, SH, VSd, ZG	52	964
4	Deutschschweiz 4: grosse Kantone (Klassenzahl > 15): SZ, SO, TG, GR, BS, AG, SG, LU	237	4595
5	Französische Schweiz (Romandie) 1 (nur drei Mindest-Jahre am Gymnasium): BEf, JU, NE, VD	152	3133
6	Französische Schweiz (Romandie) 2: FR, VSf	53	1055
7	Italienische Schweiz: Tessin	41	775
Total		747	14 593

3.2 Das Stichprobenverfahren

Die Auswahl der Maturandinnen und Maturanden erfolgte nach einem einstufigen Verfahren, bei dem Klassen proportional zu ihrer Grösse ausgewählt wurden (Cochran 1977, S. 150; Kish, 1965, S. 182; Lehtonen & Pahkinen, 1995, S. 7 [single-stage cluster sampling with probability proportional to size]). Von den ausgewählten Klassen sollten sämtliche Maturandinnen und Maturanden getestet werden. Tabelle IV.4 gibt einen Überblick über die Anzahl Klassen sowie Maturandinnen und Maturanden der Stichprobe nach Stratenbildung.

Tab. IV.4: Stichprobe

Stratum	Definition	Anzahl Klassen	Anzahl Maturandinnen und Maturanden
1	Deutschschweiz 1: Zürich	75	1439
2	Deutschschweiz 2: Bern deutsch (drei Mindest-Jahre am Gymnasium) (MD3)	30	587
3	Deutschschweiz 3: kleine Kantone (Klassenzahl <= 15): AI, NW, OW, GL, UR, AR, SH, VSd, ZG	30	578
4	Deutschschweiz 4: grosse Kantone (Klassenzahl > 15): SZ, SO, TG, GR, BS, AG, SG, LU	30	590
5	Französische Schweiz (Romandie) 1 (nur drei Mindest-Jahre am Gymnasium): BEf, JU, NE, VD	30	636
6	Französische Schweiz (Romandie) 2: FR, VSf	30	612
7	Italienische Schweiz: Tessin	35	654
Total		260	5096

3.3 Rücklaufquote

Tabelle IV.5 zeigt die Anzahl Klassen sowie Maturandinnen und Maturanden, die an der Untersuchung teilgenommen haben. Das Merkmal „teilgenommen“ wird einem Maturand beziehungsweise einer Maturandin oder einer Klasse dann zugeordnet, wenn mindestens für einen der vier durchgeführten Tests Daten vorliegen.

Tab. IV.5: Anzahl teilnehmende Klassen sowie Maturandinnen und Maturanden nach Straten

Stratum	Stichprobe Klassen	Teilnahme	Rücklaufquote Klassen	Stichprobe Personen	Teilnahme	Rücklaufquote Personen
1 Zürich	75	67	89%	1439	1204	84%
2 Deutschschweiz MD3	30	27	90%	587	496	84%
3 Deutschschweiz klein	30	25	83%	578	459	79%
4 Deutschschweiz gross	30	29	97%	590	560	95%
5 Romandie 1 (MD3)	30	27	90%	636	509	80%
6 Romandie 2	30	30	100%	612	545	89%
7 Tessin	35	10	29%	654	131	20%
Total	**260**	**215**	**90%**	**5096**	**3904**	**77%**

Nicht alle gewünschten Klassen haben an den Erhebungen teilgenommen. Problematisch ist aber einzig die aus verschiedenen Gründen niedrig ausgefallene Beteiligungsquote der Tessiner Schulen[26]. Deshalb, und weil auch die Auswahl der Klassen in einem Fall durch die Schule selbst erfolgte, also keine Zufallsauswahl mehr war, können die Ergebnisse des Tessins in diesem Teilprojekt nicht in die Auswertungen einbezogen werden. Die Rücklaufquote ohne Tessin beträgt insgesamt 91% bei den Klassen und 85% bei den Personen. Die niedrigere zweite Zahl erklärt sich aus Absenzen an den Erhebungstagen. Insgesamt lagen auswertbare Daten von rund 3’800 Personen vor.

Der Anteil der jungen Frauen beträgt in der in die Auswertungen einbezogenen Stichprobe 58.2%, jener der jungen Männer 41.8%. Diese Verhältnisse sind mit jenen der Schülerbefragung von EVAMAR I vergleichbar und spiegeln die schweizweiten Verhältnisse der Population gut wieder (vgl. Ramseier et al., 2004). Weitere Strukturmerkmale der Stichprobe werden weiter hinten im Zusammenhang mit den vergleichenden Auswertungen beschrieben.

Bei Klumpenstichproben mit ungleicher Klumpengrösse (Klassen) muss die Auswahlwahrscheinlichkeit der Klassen bei der Schätzung eines Parameters berücksichtigt werden. Weil die Klassen proportional zu ihrer Grösse ausgewählt wurden, haben nicht – wie beispielsweise bei einer einfachen Zufallsauswahl – alle Klassen beziehungsweise alle Maturandinnen und Maturanden die gleiche Auswahlwahrscheinlichkeit. Je grösser die Klasse ist, desto grösser ist ihre Wahrscheinlichkeit, in die Stichprobe zu gelangen. Aus diesem Grund mussten die Daten für die Auswertungen gewichtet werden (vgl. Satin & Shastry, 1982).

[26] So wäre wegen der tiefen Zahl der Grundpopulation faktisch eine Vollerhebung notwendig gewesen, was vor allem wegen des Erhebungszeitpunkts kurz vor den Maturaprüfungen für die Schulen eine viel grössere Belastung mit sich gebracht hätte als in den anderen Straten mit jeweils nur einzelnen ausgewählten Klassen.

4 Ablauf der Datenerhebung, Dateneingabe, Intercoder-Reliabilität

4.1 Ablauf der Datenerhebung

Die Erhebungen wurden in einem für alle Schulen vergleichbaren Zeitraum durchgeführt. Massgebend für den Erhebungszeitpunkt war die Bedingung, dass sich die Maturandinnen und Maturanden maximal drei Wochen vor Ende des regulären Unterrichts vor den Maturaprüfungen befinden sollten. Wegen der terminlichen Unterschiede der verschiedenen Schulen ergab sich daraus ein Erhebungszeitraum von Ende April bis Anfang Juli 2007. Die Testdurchführung erfolgte wie bei HarmoS durch die Schulen selbst auf der Grundlage einer genauen Anleitung. Wie in Abschnitt 3.3 bereits erwähnt, waren die Tessiner Schulen nicht in der Lage, mit der notwendigen Zahl an Maturandinnen und Maturanden teilzunehmen. Weil dieser Entscheid erst nach dem vorgesehenen Versanddatum der Materialien fiel, waren alle namentlich angeschriebenen Erhebungsunterlagen bereits versandbereit.

4.2 Codierungsverfahren bei Erstsprache

Die Antworten auf die geschlossenen Fragen wurden vor der Eingabe in die elektronischen Ergebnisdateien nicht weiter bearbeitet. Die Antworten auf die offenen Fragen hingegen mussten codiert werden. In Biologie betraf das nur wenige Aufgaben, und in Mathematik waren die Lösungen bei den offenen Fragen eindeutig und damit objektiv codierbar. Die offenen Antworten des Zusatzfragebogens wurden direkt ohne zusätzliche Codierung in die elektronische Datei übernommen.

Die Kriterien für die Codierung der offenen Antworten in Erstsprache sind weniger eindeutig als in Biologie und Mathematik. Deshalb wurde insbesondere in Erstsprache bei den Korrekturen auf eine präzise Schulung der Codierungspersonen und auf begleitende Qualitätskontrollen geachtet. Alle Korrekturaufgaben wurden nach der Erstcodierung von einer weiteren Person zweitkorrigiert. Bei Abweichungen von der Erstkorrektur wurden diese Zweifelsfälle auch in die gemeinsamen Besprechungen eingebracht. Die Testentwicklerin nahm während der ganzen Korrekturzeit Stichproben. Für die Intercoder-Reliabilitäts-Prüfung wurde eine Stichprobe aus allen Korrekturen bestimmt: parallel zur Verteilung der Schulen in den Kantonen einerseits und parallel zur ungefähren Häufigkeit der Zweitkorrekturen pro Korrektorin beziehungsweise Korrektor andererseits. Von den Korrekturaufgaben wurden jene für das Verfahren ausgewählt, die erstens inhaltlich gestuft korrigiert werden mussten (Partial-Credit-Aufgaben) und/oder zweitens nicht nur eine mögliche Lösung aufwiesen (komplexe Korrekturaufgaben). Die Intercoder-Reliabilitäts-Prüfung mit dem IRC-Koeffizienten von Früh (2004, S. 179) führte zu bei lediglich einem Item zu einem nur genügenden Wert und sonst zu guten bis hervorragenden Ergebnissen, so dass deswegen keine weiteren Items im Hinblick auf die Hauptauswertungen eliminiert werden mussten.

5 Nationale Ergebnisse

5.1 Allgemeines

Es folgt eine Darstellung der Ergebnisse der Gesamtstichprobe. Für die Testergebnisse werden dabei zwei Berechnungsvarianten vorgelegt.

Die Berechnungsvariante A dient der Veranschaulichung: Die Werte für die einzelnen Themen- und Kompetenzbereiche ergeben sich aus der Addition der Ergebnisse in den einzelnen, in die Hauptauswertung einbezogenen Aufgaben. Damit soll insbesondere der Anteil richtiger Lösungen gemessen am Total der möglichen richtigen Lösungen aufgezeigt werden. Weil ein einzelner Schüler nur zwei von mehreren Blöcken mit jeweils unterschiedlichem Umfang an Aufgaben bearbeitet hatte, wurden die Ergebnisse der Fachtests für jedes Heft pro Bereich auf jeweils maximal 20 Punkte (beziehungsweise 10 Punkte bei Biologie) standardisiert. Dass die Teilbe-

reiche in den verschiedenen Blöcken unterschiedlich schwer sind, spielt bei der Darstellung der nationalen Gesamtergebnisse insofern keine Rolle, als die Hefte nach Zufall gleich verteilt waren.

Berechnungsvariante P ist die noch genauere Schätzung der Personenfähigkeiten nach der probabilistischen Testtheorie. Probabilistische Modelle nehmen eine stochastische Beziehung zwischen dem Antwortverhalten einer Person, der Fähigkeit der Person und der Aufgabenschwierigkeit an. Diese Beziehung wird durch eine mathematische, nicht-lineare logistische Funktion grafisch dargestellt und als Item-Response-Funktion beziehungsweise als Item Response Curve (IRC) oder Item Characteristic Curve (ICC) bezeichnet (Rost, 2004, S. 96).

Abb. IV.12: Item Characteristic Curve

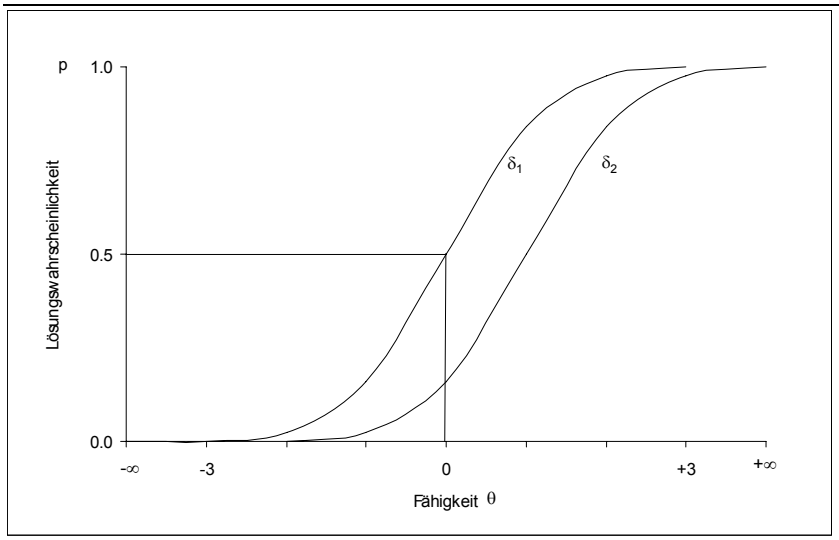


Abbildung IV.12 enthält die ICC zweier Aufgaben, die unterschiedlich schwierig sind. Aus den ICC der beiden Aufgaben mit den Schwierigkeitsparametern $\delta 1$ und $\delta 2$ lässt sich die Lösungswahrscheinlichkeit p in Abhängigkeit des Fähigkeits- beziehungsweise des Personenparameters θ ablesen (Kubinger, 2003, S. 3). Es handelt sich im vorliegenden Fall um das Rasch-Modell, bei dem sich die Wahrscheinlichkeit der richtigen Lösung einer Aufgabe als Funktion zweier Modellparameter – der Fähigkeit der antwortenden Person und der Schwierigkeit der betreffenden Aufgabe – ergibt. Je grösser die Fähigkeit ist, desto wahrscheinlicher ist es, dass eine Person eine bestimmte Aufgabe richtig löst. Und je schwieriger eine Aufgabe ist, desto unwahrscheinlicher ist es, dass eine Person mit einer bestimmten Fähigkeit die Aufgabe richtig löst. Ein grosser Vorteil dieses Modells besteht darin, dass sowohl das Testergebnis einer Person (Personenparameter) als auch die Schwierigkeiten der Aufgaben (Schwierigkeitsparameter) auf der gleichen Skala abgebildet werden. Die Skala ist so konstruiert, dass bei einer Entsprechung von Aufgabenschwierigkeit und Personenfähigkeit die Lösungswahrscheinlichkeit 50 Prozent ($p = 0{,}5$) beträgt (Rost 2004, S. 120). Die Beziehung zwischen Lösungswahrscheinlichkeit, Fähigkeit einer Person und Schwierigkeit einer Aufgabe ist für die Interpretation der Testergebnisse von entscheidender Bedeutung. Aufgrund des individuellen Testergebnisses lässt sich für jede Aufgabe bestimmen, wie wahrscheinlich es ist, dass ein Schüler oder eine Schülerin die Aufgabe richtig löst.

Ein Wert von 0 entspricht einer mittleren Personenfähigkeit. Die Skala ist grundsätzlich nach unten und nach oben offen, die Werte liegen aber in der Regel zwischen –3 und +3. Die geschätzte Personenfähigkeit stützt sich auf alle aus den verschiedenen Testheften einbezogenen Aufgaben und berücksichtigt zudem ihre unterschiedlichen Schwierigkeitsgrade.

Für die verschiedenen Vergleiche in Abschnitt 6 werden dann die auf einen Mittelwert von 500 standardisierten geschätzten Personenfähigkeiten aufgrund der Rasch-Modellierung verwendet.

Die Stichprobe wurde auf der Ebene der Klassen und nicht auf der Ebene der Maturandinnen und Maturanden gezogen (Klumpenstichprobe). Alle Maturandinnen und Maturanden einer ausgewählten Klasse wurden in die Stichprobe einbezogen. Deshalb hängt die Auswahlwahrscheinlichkeit auf Individualebene von der Grösse der

Klasse ab: Je grösser die Klasse ist, desto höher ist die Wahrscheinlichkeit, in die Stichprobe zu gelangen. Aus diesem Grund müssen die Daten gewichtet werden. Das Stichprobengewicht ergibt sich aus dem Inversen der Auswahlwahrscheinlichkeit, korrigiert nach der Antwortrate (Satin & Shastry 1982, S. 35). Das heisst, dass Fälle mit einer tieferen Auswahlwahrscheinlichkeit ein höheres Gewicht erhalten. Deshalb liegen die jeweils angegebenen Fallzahlen über der effektiven Stichprobenzahl.

Die Ergebnisse der Tessiner Erhebungen konnten, wie bereits in Abschnitt 3.3 erläutert, leider nicht einbezogen werden.

5.2 Ergebnisse in Erstsprache

5.2.1 Beschreibung der Gesamtergebnisse

In Tabelle IV.6 sind die Gesamtergebnisse in Erstsprache aufgeführt. Der aus konstruktionstechnischen Gründen nur in der Deutschschweiz eingesetzte Teil „Grammatische Kompetenz und Orthografie“ des Erstsprachtests ist im Gesamttotal nicht enthalten und wird separat ausgewiesen. Bei der Berechnungsvariante P ermöglicht die Verlinkung der Testhefte Fähigkeitsschätzungen für diesen Bereich auch für die Romandie.

Tab. IV.6: Gesamtergebnisse Erstsprache

Bereiche	Mittel A	S A	Max A	Mittel P	S P	Min P	Max P
TOT_SPRACHE	31,22	9,27	60,00	,20	,562	–1,84	2,03
ALLG_LV	9,89	4,94	20,00	,38	,621	–2,11	2,30
DETAIL_LV	10,52	3,97	20,00	,44	,676	–2,29	2,31
WORTS	10,81	4,39	20,00	,39	,424	–1,18	1,90
GRAM	13,49	2,60	18,92	,38	,399	–1,36	2,32

N = 13 361 (gewichtet)

Die Abkürzungen in der Tabelle verstehen sich wie folgt:

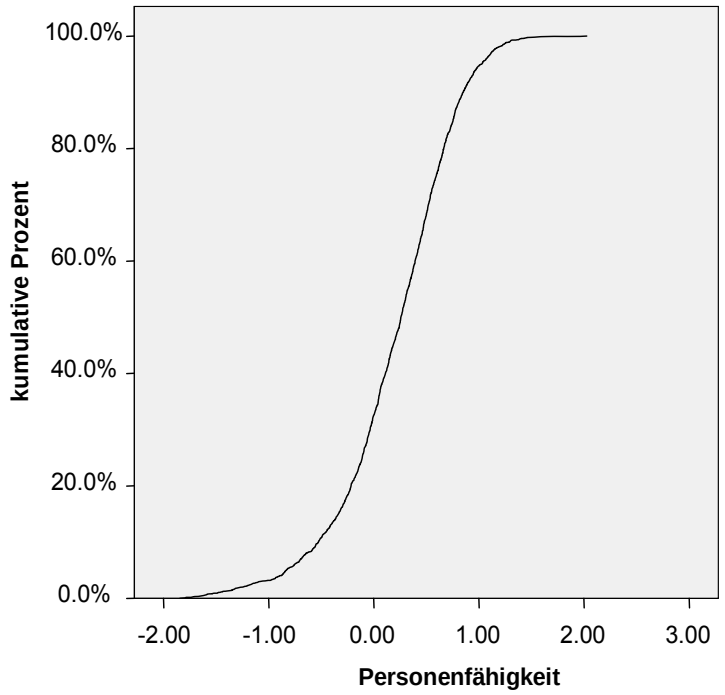
- TOT_SPRACHE: Gesamtergebnis Erstsprache (ohne Grammatik)
- ALLG_LV: Allgemeines Leseverstehen (zur Orientierung lesen)
- DETAIL_LV: Detailliertes Leseverstehen (Information und Argumentation verstehen)
- WORTS: Sprachreflexion/Wortschatz
- GRAM: Sprachreflexion/Grammatische Kompetenz und Orthografie
- Mittel A: Mittelwerte nach Variante A (max. 20 Punkte pro Bereich, max. 60 Punkte total)
- S A: Standardabweichung zu Mittel A
- Max A: Erreichte Maximalpunktzahl Variante A
- Mittel P: Mittelwerte nach Variante P (Wert null entspricht einer „mittleren“ Fähigkeit)
- S P: Standardabweichung zu Mittel P
- Min P: Tiefste Personenfähigkeit Variante P
- Max P: Höchste Personenfähigkeit Variante P

Im Mittel wurden die Aufgaben zu etwas mehr als der Hälfte richtig gelöst (im Bereich „Grammatische Kompetenz und Orthografie“ deutlich mehr). Die geschätzten Personenfähigkeiten liegen im Durchschnitt in allen Kompetenzbereichen auf einem Niveau, auf dem die Maturandinnen und Maturanden Aufgaben lösen konnten,

deren Anforderungen über einem mittleren Schwierigkeitsgrad liegen. Die Streuung auf der Ebene der einzelnen Personen ist beachtlich. Mit Ausnahme des Bereichs „Grammatische Kompetenz und Orthografie“ wurde jeweils auch die maximal mögliche Punktzahl erreicht. Eine zusätzliche Auswertung der über die Klassen aggregierten Werte ergibt eine erhebliche Streuung auch zwischen den Klassenmittelwerten (Minima: TOT_SPRACHE = 20.72, GRAM = 7.55; Maxima: TOT_SPRACHE = 41.41, GRAM = 17.23). Die beste Klasse hat im Mittel doppelt so viele Aufgaben richtig gelöst wie die schlechteste.

In Abbildung IV.13 ist die kumulative Verteilung der Ergebnisse für die Gesamtschätzung der Personenfähigkeiten in Erstsprache grafisch dargestellt.

Abb. IV.13: Ergebnisse Personenfähigkeiten Erstsprache Total



5.2.2 Beschreibung der Ergebnisse ausgewählter Items

In diesem Abschnitt werden ausgewählte Items beschrieben, die unterschiedlich gut gelöst wurden. Einerseits werden die effektiven, relativen Lösungshäufigkeiten angegeben, andererseits auch die Rasch-geschätzten objektiven Aufgabenschwierigkeiten δ. Eine Person mit der nach Berechnungsvariante P ermittelten, gleichen Personenfähigkeit würde eine solche Aufgabe mit einer Wahrscheinlichkeit von 50% richtig lösen (siehe auch die Erläuterungen in Abschnitt 5.1). Ein Wert von 0 entspricht einer mittleren Aufgabenschwierigkeit. Die Skala ist grundsätzlich nach unten und nach oben offen, die Werte liegen aber in der Regel zwischen –3 und +3.

Ein erstes Beispiel ist Aufgabe 4.7 (siehe Abbildung IV.14) aus dem Bereich „Wortschatz“, die von 77.3% der Maturandinnen und Maturanden richtig gelöst wurde (δ = -0.98). Hier geht es um die Kenntnis von Fremdwörtern im Bereich Adjektive und um Antonyme.

Abb. IV.14: Beispiel einer leichten Aufgabe zum Wortschatz

Aufgabe 4.7 **Welches Adjektiv ersetzt *„intendiert"* im folgenden Satzteil?** **Kreuzen Sie die richtige Antwort an.** „Die individuellen und sozialen, intendierten und zufälligen, kurz- wie langfristigen, sozial erwünschten, aber auch schädlichen Effekte der Massenmedien auf Wis-sen, Einstellungen, Emotionen und Verhaltensweisen ..." (Bonfadelli [2005], S. 12) 1 ❑ erwirkten 2 ❑ erzeugten 3 ❑ verstrickten 4 ❑ beabsichtigten	D_W_47a

Eine Analyse der Distraktoren zeigt, dass im Gegensatz zur leicht als falsch erkennbaren Option 1 „erwirkten" (3.1% der Nennungen) und Option 3 „verstrickten" (3.8% der Nennungen) der schwierigere Distraktor 2 „erzeugten" mit 12.6% der Nennungen relativ häufig gewählt wurde. Hier befanden sich die Maturandinnen und Maturanden mit dem gewählten antonymischen Adjektiv auf einem Lösungsweg, auf dem sie versuchten, das ihnen anscheinend nicht bekannte Fremdwort in das erkannte Gegensatzpaar-Muster einzupassen. „Erzeugt" als Gegensatz zu „zufällig" erscheint in dieser gewählten Übersetzungsstrategie sogar passend.

Aus dem Bereich „Detailliertes Leseverstehen/Informationen und Argumentation verstehen" stammt das Item in Abbildung IV.15, das entgegen der Experteneinschätzung aufgrund der erbrachten Maturandenleistungen als sehr leicht eingestuft werden kann (δ = -1.96). Es handelt sich um eine geforderte Richtigstellung eines Fehlersatzes, der auf einem Primärtext einer Einführung in das Studienfach Geschichte beruht. Hier ist ein genaues Nachvollziehen der differenten Informationsstränge des Einstiegstextes (Zeile 1–9: Rolle der beiden Gerichtshöfe, Verhältnis der Hinrichtungen in Bezug auf die Verurteilungen, Tötungsart) gefordert, und im zweiten Schritt die Fähigkeit, selektiv die wichtigsten Unterschiede zwischen den Distraktoren (verschiedene Tötungsarten) zu fokussieren und die restlichen Zusatzinformationen zu vernachlässigen. Diese Aufgabe lösten 89.8% aller getesteten Maturandinnen und Maturanden richtig.

Abb. IV.15: Beispiel einer sehr leichten Aufgabe des detaillierten Leseverstehens

Textauszug 5.1 Lesen Sie den Text und beantworten Sie daraufhin die folgenden Fragen. „In der Frühen Neuzeit, also etwa zwischen 1450 und 1750, wurden in Europa Tausende von Menschen, vorwiegend Frauen, wegen des Verbrechens der Hexerei vor Gericht gestellt. Etwa die Hälfte von ihnen wurde hingerichtet, meistens auf dem Scheiterhaufen verbrannt. Einige Hexenprozesse wurden von den verschiedenen kirchlichen Gerichtshöfen Europas durchgeführt, Institutionen, die im Mittelalter und in der Frühen Neuzeit bei der Überwachung von Moral und Religion eine wichtige Rolle spielten. Häufiger jedoch und besonders nach 1550 führten weltliche Gerichtshöfe, die Gerichte der Königreiche, Staaten (...) und Städte die Verfahren. (…) (Levack, B.P. [2003]. *Hexenjagd. Die Geschichte der Hexenverfolgungen in Europa*, S. 11)	
Aufgabe 5.2 Der folgende Satz enthält Fehler. Vergleichen Sie ihn mit dem Eingangstext von 5.1 und stellen Sie die Aussage richtig. Kreuzen Sie die richtige Antwort an. Im Unterschied zu den weltlichen Gerichtshöfen verurteilten die kirchlichen Gerichtshöfe die meisten der angeklagten Frauen zum Tod durch den Scheiterhaufen. 1 ❑ Die kirchlichen Gerichtshöfe verurteilten die angeklagten Frauen meistens zur „Wasserprobe“, dem Tod durch Ertrinken. 2 ❑ Die weltlichen Gerichtshöfe verurteilten keine Frauen zum Tode. 3 ❑ Sowohl die kirchlichen als auch die weltlichen Gerichtshöfe verurteilten Angeklagte zum Tod durch den Scheiterhaufen. 4 ❑ Wie die weltlichen Gerichtshöfe verurteilten auch die kirchlichen Gerichtshöfe die angeklagten Frauen zum Hungertod im Turm.	D_D_52a

Sehr leichte Items fanden sich auffallend häufig im Grammatik-Teil der Aufgaben in der Erstsprache Deutsch. Folgendes Beispiel vermochten 72.2% aller getesteten Maturandinnen und Maturanden richtig zu lösen (siehe Abbildung IV.16) (δ = -0,57).

Abb. IV.16: Beispiel einer leichten Grammatikaufgabe

Textauszug 8.1 (…) Ein weiteres Beispiel für eine Theorie ist die Rezeptortheorie. Die Zuckerkrankheit, der Diabetes mellitus, ist eine schon seit langem bekannte Erkrankung und wurde früher als Ausscheidung honigsüssen Urins bezeichnet. Frerichs, ein Berliner Pathologe, beobachtete im vorletzten Jahrhundert bei Patienten mit einem Diabetes mellitus, dass die Bauchspeicheldrüse dieser Patienten unter dem Mikroskop anders aussah als bei Gesunden und damit der Diabetes eine Erkrankung der Bauchspeicheldrüse ist. Claude Bernard, der berühmte französische Physiologe und Gründervater der experimentellen Medizin, wies nach, dass Zucker in der Leber produziert wird, und so zählte man den Diabetes zu den Lebererkrankungen. (…) (Steurer [2005]. *Wissenschaftstheoretische Grundlagen der Medizin*, Skript zur Vorlesung, OLAT-pdf, S. 7–9)	
Aufgabe 8.6 Sie finden in oben stehendem Text 8.1 mehrere Sätze, in denen „Diabetes“ vorkommt. Welches grammatikalische Geschlecht hat dort dieses Wort? Kreuzen Sie die richtige Antwort an. 1 ❑ femininum 2 ❑ neutrum 3 ❑ maskulinum 4 ❑ keines	D_G_86

Die Aufgabenstellung erfordert kein Vorwissen, das auf genauen Wortschatzkenntnissen beruht, sondern kann bei genauem Lesen relativ einfach aus dem beigefügten Textausschnitt herausgefiltert werden, da das gefragte Wort gleich viermal wiederholt wird (Zeilen 2, 5, 7, 10). Die Aufgabe stellt somit keine hohen Ansprüche an Maturandinnen und Maturanden, sondern sollte nach Einschätzung der Testkonstrukteure von 100% der Probandinnen und Probanden richtig gelöst werden.

Eine Betrachtung der Häufigkeitswerte der Distraktoren zeigt, dass 10.5% aller Probandinnen und Probanden die Option mit dem femininen und 6.8% jene mit dem sächlichen Artikel wählten.

Zur Beschreibung von Aufgaben, die vielen schweizerischen Gymnasiastinnen und Gymnasiasten schwer bis sehr schwer fielen, dient das exemplarische Beispiel in Abbildung IV.17 aus dem Bereich des allgemeinen Leseverständnisses:

Abb. IV.17: Beispiel einer schwierigen Textverständnis-Aufgabe (allgemeines Leseverständnis)

Text 4.1

(…)

„Die universitäre Verankerung und die Herausbildung der Identität der verhältnismässig jungen sozialwissenschaftlichen Disziplin Publizistik- und Kommunikationswissenschaft und die Bestimmung ihres Gegenstandes sind sowohl im deutschen als auch im englischen Sprachraum mit Schwierigkeiten und Spannungen verbunden (gewesen). Sichtbar wird dies in der Vielzahl der bestehenden und sich konkurrenzierenden Fachbezeichnungen wie Publizistik-, Kommunikations- sowie Medienwissenschaft oder gar Journalistik, aber auch in den verschiedensten neuen Handbüchern und Lehrbüchern in diesem Bereich (vgl. u. a. Krallmann/Ziemann 2001; Kunczik 2001 etc.). Trotz dieser Schwierigkeiten geht die Deutsche Gesellschaft für Publizistik- und Kommunikationswissenschaft (DGPuK) davon aus, dass es sich um ein Fach handelt, das sich mit ähnlichen Problemen und Gegenständen beschäftigt, nämlich der öffentlichen Kommunikation.

Aus der Aussenperspektive erscheint das Fach jedoch als heterogen, bestenfalls als *Bindestrich-Wissenschaft*. Praktisch alle Disziplinen der Geistes- und Sozialwissenschaften beschäftigen sich nämlich, allerdings aus ihrer jeweiligen Fachperspektive heraus, mit Teilbereichen des Gegenstandes (Massen-)
Kommunikation wie Medienphilosophie, Mediengeschichte, Medienökonomie, Medienpolitik, Medienrecht, Mediensoziologie, Medienpsychologie, Medienpädagogik etc., um nur einige zu nennen."

(Bonfadelli, H., Jarren, O., Siegert, G. [Hrsg.], [2005]. *Einführung in die Publizistikwissenschaft*. Bern:Verlag Haupt. S. 5)

Aufgabe 4.3

Folgende Hinweise (Marginalien) sind im Lehrbuch zur Orientierungshilfe seitlich am Rande des Textes 4.1 notiert.

Wie ist ihre richtige Reihenfolge?

1 Fachverband
2 nicht homogen
3 Namensvielfalt
4 unterschiedliche Sichtweisen
5 Entstehung des Faches

Die richtige Reihenfolge der Hinweise ist:

(Schreiben Sie die Zahl in ein Kästchen)

☐ ☐ ☐ ☐ ☐

D_A_43

Diese Aufgabe 4.3, welche auf einem relativ dichten Einführungstext in die Kommunikations- und Publizistikwissenschaft beruht, erfordert allgemeines Leseverstehen in dem Sinne, als es zum Lösen der Aufgabenstellung notwendig ist, sich einen Überblick über den Handlungsverlauf oder Erzählstrang des Textes zu verschaffen. Es müssen dazu die Hauptthemen der einzelnen Abschnitte erfasst werden; für die Aufgabenstellung unwesentliche Detail-Informationen können hingegen vernachlässigt werden (z. B. Zeile 11: die volle Bedeutung von

DGPuK). Spezifisch an dieser Aufgabe ist, dass die Themen der einzelnen (Unter-)Abschnitte in eine sequenzielle Abfolge zu bringen sind; das heisst, die Maturandinnen und Maturanden müssen ihr Textverständnis entlang dem linearen Textverlauf entwickeln. Dieses sequenzielle Lesen stellt eine Grundlage des genauen und vollständigen, nicht-selektiven Lesens dar und ist somit eine Basiskompetenz (Kintsch, 1998; Gernsbacher, 1997). Die vorgestellte Aufgabe 4.3 wurde gemäss der Einschätzung der Testkonstrukteurin als eher leicht bis mittel eingestuft, da die Hauptthemen des Textinhaltes bereits vorgegeben sind und nicht selber eruiert werden müssen. Die eigentliche Schwierigkeit besteht in der Verknüpfung der Hinweise mit den passenden Textstellen. So muss beispielsweise die Marginalie „Namensvielfalt" verbunden werden mit den Zeilen 5–7: „Sichtbar wird dies in der Vielzahl der bestehenden und sich konkurrenzierenden Fachbezeichnungen wie Publizistik-, ... sowie Medienwissenschaft oder gar Journalistik, ..." oder das Signalwort „heterogen" (Zeile 14) mit der vorgegebenen Randbemerkung „nicht homogen". Sind erste Brücken einmal gebildet, können die anderen Zuordnungen umso leichter erschlossen werden. Erstaunlicherweise konnte diese Aufgabe des allgemeinen Leseverstehens nur von rund 21.9% aller getesteten Maturandinnen und Maturanden richtig gelöst werden (δ = 1.77). Eine detaillierte Analyse der Häufungen von falschen Reihenfolgen könnte allenfalls weitere Informationen bringen.

Zwei weitere schwierige bis sehr schwierige Aufgaben aus dem Bereich des allgemeinen Leseverstehens unterscheiden sich von obigem halboffenen Beispiel erstens durch das deutlich offene Format und zweitens durch das Codierungsverfahren des Partial-Credit (abgestufte Punktebewertung) (siehe Abbildung IV.18 und IV.19).

Aufgabe 8.1 besteht darin, für einen in drei Abschnitte gegliederten, mittellangen Vorlesungstext einen angemessenen Titel zu finden; es handelt sich inhaltlich um eine relativ einfach verständliche Einführung in die Medizingeschichte mit einer knappen Übersicht von Hippokrates bis zur Diabetes-Forschung des 20. Jahrhunderts. Die Aufgabenstellung erwies sich als die drittschwierigste von den in der Hauptauswertung verbliebenen offenen Aufgaben. Sie wurde mit einer fünfstufigen Skala korrigiert von 1 = halbrichtig[27], 2 = richtig (einfach), 3 = komplex richtig, 4 = mit Zusatzwissen bis 5 = mit kreativer Sprachleistung. Am häufigsten wurde die halbrichtige Stufe 1 codiert (39.2%), gefolgt von den einfachen richtigen Lösungen auf Stufe 2 (18.8%). Nur 2% der Maturandinnen und Maturanden haben bei dieser Titelsetzungsaufgabe komplexere Lösungen auf der Stufe 3 formuliert, nur 1.3% zeigten Zusatzwissen (Stufe 4) und nur 2.4% kreative oder sonstwie herausragende Sprachleistungen (Stufe 5). Die Rasch-geschätzten objektiven Schwierigkeiten δ dieser Aufgabe lauten für die Stufen 1 bis 5: -0.02, 1.18, 1.69, 1.78, 1.94.

[27] Als halbrichtig waren Lösungen definiert, die wichtige Hauptaspekte des Textes vernachlässigten, neben einer richtigen Aussage auch falsche Elemente enthielten oder so allgemein gehalten waren, „dass sie zwar nicht falsch sind, aber das Spezifische des Textes nicht wiederzugeben vermögen" (Korrekturschlüssel, S. 7), beispielsweise „Die Geschichte menschlicher Krankheiten" oder „Irrtümer der Medizin."

Abb. IV.18: Beispiel einer sehr schwierigen Textverständnis-Aufgabe (allgemeines Leseverständnis) im offenen Format (fünfstufige Partial-Credit-Codierung)

Aufgabe 8.1 Was für einen Titel könnte die folgende Abhandlung tragen? „Der Mensch, so die Theorie des Hippokrates, besteht im Wesentlichen aus vier Säften: Diese sind das Blut, das Phlegma und die schwarze und gelbe Galle. Wenn diese Säfte im Gleichgewicht sind, so ist der Mensch gesund, und wenn diese Säfte in ein Ungleichgewicht geraten, so ist der Mensch krank. Wenn jemand Fieber bekam, und das ist, wie wir heutzutage wissen, sehr oft bei Infektionen der Fall, so hatte der Mensch nach der hippokratischen Theorie zu viel Blut. Die logische Konsequenz daraus, um den Patienten vom Fieber zu kurieren, ist die Blutmenge zu verringern, also den Patienten zur Ader zu lassen. Diese Theorie hielt sich sehr lange und erst Robert Koch, ein Forscher in Berlin, fand heraus, dass unter anderem Bakterien und nicht zuviel Blut die Ursache von Infektionskrankheiten sind. Eine Infektion, zum Beispiel die Tuberkulose ist also nicht mit einem Ungleichgewicht der Körpersäfte zu erklären, sondern durch einen Erreger, der in den menschlichen Körper eindringt und dort zu Veränderungen in der Lunge und in der Folge zu Fieber führt. Diese Theorie wird die Erregertheorie genannt. Ein weiteres Beispiel für eine Theorie ist die Rezeptortheorie. Die Zuckerkrankheit, der Diabetes mellitus, ist eine schon seit langem bekannte Erkrankung und wurde früher als Ausscheidung honigsüssen Urins bezeichnet. Frerichs, ein Berliner Pathologe, beobachtete im vorletzten Jahrhundert bei Patienten mit einem Diabetes mellitus, dass die Bauchspeicheldrüse dieser Patienten unter dem Mikroskop anders aussah als bei Gesunden und damit der Diabetes eine Erkrankung der Bauchspeicheldrüse ist. Claude Bernard, der berühmte französische Physiologe und Gründervater der experimentellen Medizin, wies nach, dass Zucker in der Leber produziert wird, und so zählte man den Diabetes zu den Lebererkrankungen. Langerhans hat 1869 die später nach ihm benannten Inselzellen in der Bauchspeicheldrüse erstmals beschrieben und zusammen mit anderen Forschern die Hypothese aufgestellt, dass diese Zellen den Zuckerstoffwechsel regulieren.(...) Zwei Forschern in Kanada, sie hiessen Bantig und Best, gelang 1921 die Isolierung der Inselzellen. (...) Sie injizierten ein Extrakt von Inselzellen einem an der Zuckerkrankheit verendenden Hund. (...) Nach Selbstexperimenten injizierten sie diese Extrakte erfolgreich „zuckerkranken" Patienten, und in der Folge wurde Insulin für viele Menschen eine lebensrettende Therapie. Damit wusste man nun, dass Insulin den Zuckergehalt im Blut regelt." (Steurer [2005]. *Wissenschaftstheoretische Grundlagen der Medizin*, Skript zur Vorlesung, OLAT-pdf, S. 7–9)	D_A_81

Bei der anschliessenden Aufgabe 8.2 (Abbildung IV.19), die im selben Text das Setzen von Zwischentiteln für die drei inhaltlich differierenden Abschnitte verlangt, zeigt sich – obwohl sie nur im vierstufigen Partial-Credit-Verfahren korrigiert wurde – ein ähnliches Grundmuster in der Häufigkeitsverteilung der Skalenstufen.

Abb. IV.19: Beispiel einer schwierigen Textverständnis-Aufgabe (allgemeines Leseverständnis) im offenen Format (vierstufige Partial-Credit-Codierung)

Aufgabe 8.2 Finden Sie für jeden Abschnitt von Text 8.1 einen passenden Zwischentitel!	
a	D_A_82a
.....	
b	D_A_82b
.....	
c	D_A_82c
.....	

Diese drei Teilaufgaben wurden erwartungsgemäss besser gelöst als die vorangehende, die eine Übersicht über die Inhalte aller drei Abschnitte und deren Verknüpfung erfordert. Die Rasch-geschätzten objektiven Schwierigkeiten δ dieser Aufgaben lauten für die Stufen 1 bis 4 wie folgt: Aufgabe 8.2a: -1.12, 1.07, 1.96, 2.18; Aufgabe 8.2b: -1.85, 1.77, 2.51, 2.59; Aufgabe 8.2c: -1.49, 1.25, 2.38, 2.45. Der erste und dritte Abschnitt, die länger und inhaltlich komplexer sind als der mittlere, weisen beide eine beinahe gleiche Verteilung der effektiven Lösungshäufigkeiten auf: 52.4% beziehungsweise 58.3% der richtigen Titel befinden sich auf der einfachen Stufe 1, nur 25.7% beziehungsweise 27.7% auf der komplexeren Stufe 2. Bei dem sehr kurzen zweiten Abschnitt (nur vier Zeilen) setzten sogar 72.4% aller Probandinnen und Probanden einen einfachen Titel, meist ein Wort (z. B. „Erreger“, „Erregertheorie“, „Tuberkulose“). Komplexere Zwischentitel der Stufe 2 formulierten 16.5% (z. B. „Erklärung einer Infektion – am Beispiel von Tuberkulose“). Wie bei der Gesamttitelsetzung zeigten bei allen drei Zwischentiteln bloss 0.4% bis 3% Zusatzwissen und auch nur 0.9% bis 3.2% kreative oder sonstwie herausragende Sprachleistungen.

Bemerkenswert sind auch die Schüleranteile, die nicht in der Lage waren, die Basisanforderungen an einen treffenden Titel zu erfüllen: Knapp ein Viertel der Maturandinnen und Maturanden (23.6%) setzten bei dem Gesamttitel von Aufgabe 8.1 einen falschen[28] oder nicht genügenden Titel. Bei den Zwischentiteln bewegen sich die Anteile inhaltlich falscher oder inhaltlich nicht zum Abschnitt passender Formulierungen zwischen 13.7% beim Abschnitt über die Antike (z. B. „Hippokratische Theorie der Infektionstheorie“) und 7.8% beim kürzesten Abschnitt von vier Zeilen (z. B. „Erklärungen“ oder „gesund oder krank?“).

Als Beispiel eines mittelschweren Items sei im Folgenden eine Aufgabe (Abbildung IV.20) vorgestellt, die von etwa der Hälfte aller getesteten Maturandinnen und Maturanden richtig gelöst wurde (54%) (δ = 0.27).

[28] Die Lösung für den Gesamttitel wurde in den folgenden Fällen mit 0 Punkten bewertet: Der Titel gibt die Gesamtaussage des Textes falsch wieder, weil er „a) sich nur auf einen Abschnitt bezieht, b) nicht die Aussage des Textes wiedergibt, c) nur einen von den vier wichtigen Aspekten erwähnt oder d) zwar zwei von den vier Aspekten [Evolution/Medizin/Theorie & Forschung/Therapie] erwähnt, aber ein zusätzliches falsches Element dazu kombiniert, oder es sich um einen falschen Bezug handelt“ (Korrekturschlüssel, S. 7).

Beispiele zu Code 0 der Zwischentitel sind „Widerlegte Theorien“, „Der Mensch wird von Hormonen gesteuert“, „Geschichte der Infektionen“.

Abb. IV.20: Beispiel einer mittelschweren Wortschatz-Aufgabe

Textauszug 6.1 (...) f) Pilze färben ganze Landstriche! Wer hätte sich nicht schon über knallbunte Fruchtkörper von Schlauch- und Ständerpilzen gefreut (z. B. Becherlinge, Fliegenpilz, Saftlinge etc.) Insbesondere Flechtenbildner prägen (...) oft ganze Gegenden, z. B. safrangelbe Küstenfelsen, Hausdächer, Bäume dank Anthrachinonen in *Xanthoria*-Arten (...).Viele dieser Flechten-(...) sind autofluoreszent: sie absorbieren UV und emittieren längerwelliges Licht." (Honegger, Rosmarie [2005]. *Pilze (Mykologie)*. Skript SK-Nr. 741. Zürich: Universität. S. 9)	
Aufgabe 6.2 **Was bedeutet im Text 6.1.f *„autofluoreszent"*?** **Kreuzen Sie die richtige Antwort an.** 1 ❑ selber Fluor abgebend 2 ❑ selbst ablagernd 3 ❑ selbst(auf)leuchtend 4 ❑ reflektierend	D_W_62

Die Distraktoren-Analyse zeigt, dass die Option 4 „reflektierend" von 31.1% aller Testpersonen gewählt wurde. Dies kann als Indikator dafür gesehen werden, dass diese Probandinnen und Probanden zusätzlich zum Fremdwort-Adjektiv „autofluoreszent" auch die Fremdwort-Verben „absorbieren" und „emittieren" nicht kennen, da der Satzinhalt durch das gewählte Einsetzen von Distraktor 4 unlogisch wird (werden UV-Strahlen aufgenommen, können sie nicht gleichzeitig zurückgeworfen werden).

5.3 Ergebnisse in Mathematik

5.3.1 Beschreibung der Gesamtergebnisse

In Tabelle IV.7 sind die Gesamtergebnisse in Mathematik aufgeführt.

Tab. IV.7: Gesamtergebnisse Mathematik

Bereiche	Mittel A	S A	Max A	Mittel P	S P	Min P	Max P
TOT_MATH	37.65	13.25	80.00	–.09	.721	–2.22	2.44
ANALYSIS	7.79	4.71	20.00	–.12	.896	–2.70	2.90
GEOMETRIE	9.08	4.58	20.00	–.08	.774	–2.28	2.49
STOCHASTIK	11.02	5.12	20.00	–.08	.990	–3.46	3.24
M_VERSCH	9.76	4.76	20.00	–.11	.818	–2.55	2.64

N = 13'524 (gewichtet)

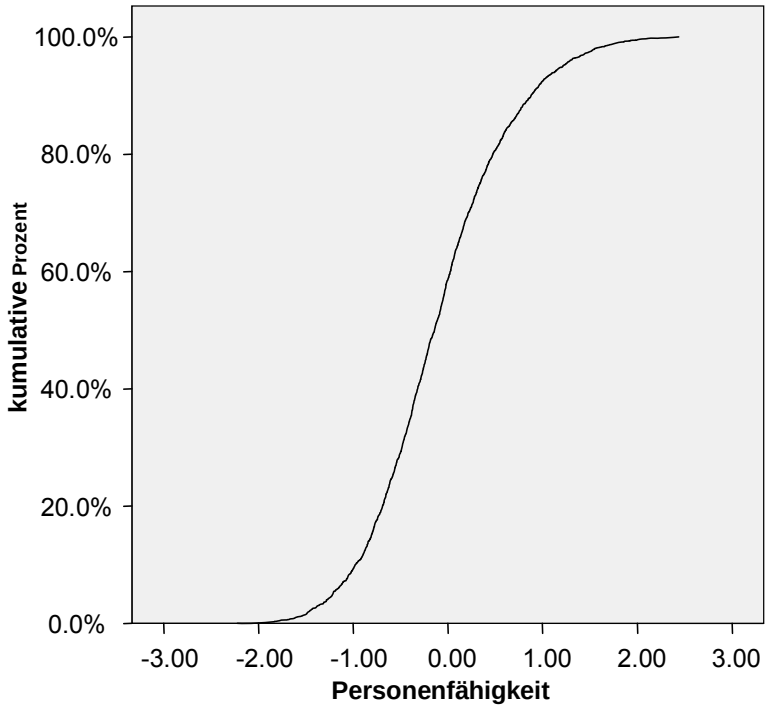
Die neu verwendeten Abkürzungen in der Tabelle bedeuten Folgendes:

- TOT_MATH: Gesamtergebnis Mathematik
- ANALYSIS: Analytisches (Funktionen, Infinitesimalrechnung, Folgen und Reihen)
- GEOMETRIE: Geometrisches (Trigonometrie, analytische Geometrie, Vektorrechnung)
- STOCHASTIK: Stochastisches (Wahrscheinlichkeitsrechnung, Kombinatorik, Statistik)
- M_VERSCH: Diverses (Mengenlehre/diskrete Mathematik, Logik, elementare Algebra)
- Mittel A: Mittelwerte nach Variante A (max. 20 Punkte pro Bereich, max. 80 Punkte total)

Im Mittel wurden die Aufgaben von den Testpersonen zu etwas weniger als der Hälfte richtig gelöst, die Analysisaufgaben zu rund 40%, die Geometrieaufgaben zu 45%, die Stochastikaufgaben zu 55% und die Aufgaben des übrigen Bereichs zu 50%. Werden die Aufgabenschwierigkeiten in die Schätzung der Personenfähigkeiten einbezogen, nähern sich die Durchschnittswerte einander an, insbesondere bei den Bereichen Analysis (gegen oben) und Stochastik (gegen unten). Das eher schlechte Rohergebnis in Analysis muss deshalb insofern relativiert werden, als diese Aufgaben besonders schwierig waren. Insgesamt liegen die geschätzten Personenfähigkeiten in allen Kompetenzbereichen im Durchschnitt jeweils auf einem Niveau, auf dem die Maturandinnen und Maturanden Aufgaben lösen konnten, deren Anforderungen ganz leicht unter einem mittleren Schwierigkeitsgrad liegen. Dieser Wert liegt unter den mittleren Personenfähigkeiten in Erstsprache und über jenen in Biologie für diejenigen Maturandinnen und Maturanden, die Biologie bereits abgeschlossen hatten. Die Streuung zwischen den einzelnen Personen ist wieder erheblich. In allen Bereichen wurde auch das Punktemaximum erreicht. Eine zusätzliche Auswertung der über die Klassen aggregierten Werte ergibt erneut eine bemerkenswerte Streuung auch zwischen den Klassenmittelwerten (Minimum: TOT_MATH = 22.36; Maximum: TOT_MATH = 64.04). Die beste Klasse hat also im Mittel beinahe dreimal so viele Aufgaben richtig gelöst wie die schlechteste.

In Abbildung IV.21 ist die kumulative Verteilung der Ergebnisse für die Gesamtschätzung der Personenfähigkeiten grafisch dargestellt.

Abb. IV.21: Ergebnisse Personenfähigkeiten Mathematik Total



Eine detaillierte Analyse der Lösungshäufigkeiten der einzelnen Testfragen erlaubt die folgenden, wegen der unterschiedlichen Schwierigkeiten der Aufgaben vorsichtig zu lesenden Aussagen: Relativ gut umgehen konnten die Testpersonen mit einigen Fragestellungen wie der Auswertung von Funktionstermen mit zwei Variablen, der Auswertung von Summen mit mehreren Indizes und Fragen zur Logik (Wahrheitstafeln). Besonders hoch sind die Werte auch in der Vektorrechnung und in der Kombinatorik. In der Logik gibt es aber auch Aufgaben, die besonders schlecht gelöst wurden (siehe weiter unten in Abschnitt 5.3.2). Relativ schlecht gelöst wurden Aufgaben zu Exponential- oder Logarithmusfunktionen. Die Bearbeitung von Fragen zu Folgen und Reihen erwies sich ebenfalls als nur unterdurchschnittlich. Das könnte eine Folge der vermutlich abnehmenden Bedeutung dieses Gebiets an den Gymnasien sein. Auch Fragen zu Gleichungssystemen und Fragen zu quadratischen Gleichungen wurden nur unterdurchschnittlich gut bewältigt.

Bezogen auf die Fähigkeitsdimension des Kompetenzrasters fällt auf, dass Aufgaben, die Ansprüche an das Textverständnis stellen, sehr gut gelöst wurden. Das Rechnen mit Zahlen ist mit überdurchschnittlichen Lösungshäufigkeiten viel besser bewältigt worden als das Rechnen und Operieren mit allgemeinen Termen, die in der Regel Variablen und/oder Unbekannte enthalten. Beim Interpretieren von Diagrammen sind die Lösungshäufigkeiten im Mittel ebenfalls recht gut.

5.3.2 Beschreibung der Ergebnisse ausgewählter Items

In diesem Abschnitt werden einige Testaufgaben des Mathematiktests beschrieben, die unterschiedlich gut gelöst wurden. Wie schon beim Erstsprache-Test werden die effektiven, relativen Lösungshäufigkeiten sowie die Rasch-geschätzten objektiven Aufgabenschwierigkeiten δ angegeben.

Die leichteste Aufgabe M_X_10 bestand im Ausfüllen einer Wahrheitstafel, war also eine Aufgabe aus dem Gebiet der Logik (siehe Abbildung IV.22). Da solche Tafeln vermutlich im Unterricht nicht (mehr) behandelt werden, ist Textverständnis wichtig. Die Aufgabe war in drei Teilaufgaben unterteilt. Jede Spalte der auszufüllenden Tafel bildet eine Teilaufgabe. 91% der Testpersonen haben die Spalte „A“ richtig ausgefüllt (δ = -2.59), 84% die Spalte „A oder B“ (δ = -0.98) und 83% die Spalte „B und C“ (δ = -1.89). 91% richtige Lösungen ist der höchste beim Mathematiktest erreichte Wert.

Abb. IV.22: Beispiel 1 einer gut gelösten Mathematikaufgabe

M_X_10

Hier ist ein Auszug aus einer Datenbank von Studierenden.

Nummer	Name	Geschlecht	Nationalität	Fakultät
1	Anna	weiblich	Frankreich	Architektur
2	Sergio	männlich	Italien	Wirtschaft
3	Miriam	weiblich	Schweiz	Kommunikation
4	Luc	männlich	Frankreich	Wirtschaft

Nun bedeute

A: "studiert Wirtschaft"

B: "hat Nationalität Schweiz"

C: "ist weiblich"

Vervollständigen Sie die folgende Wahrheitstafel.

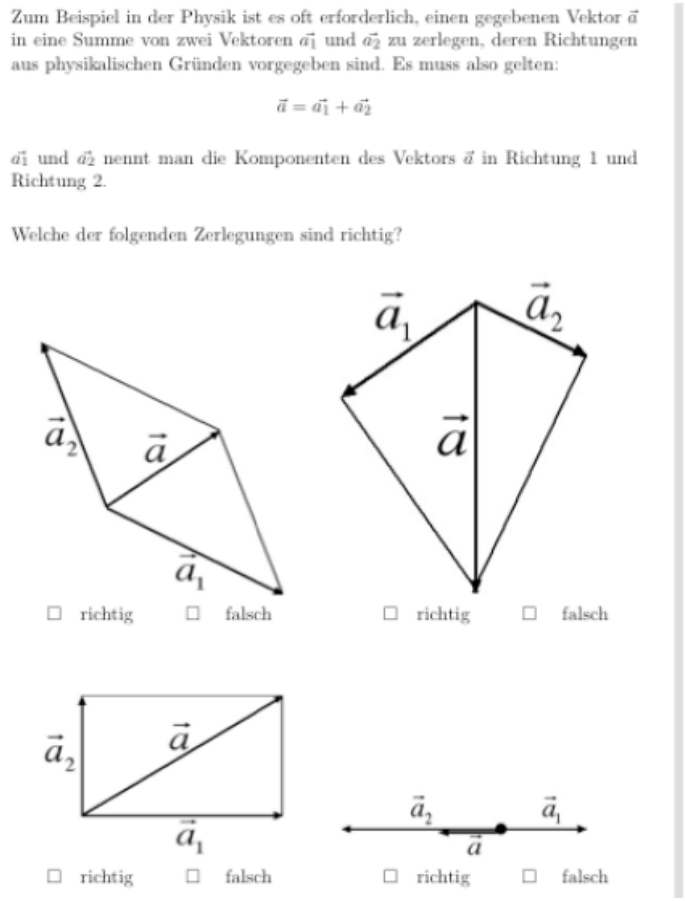
Nummer	Name	A	A oder B	B und C
1	Anna	f		
2	Sergio	w		
3	Miriam			
4	Luc			

w bedeute wahr, f bedeutet falsch

Ebenfalls sehr gut gelöst wurde die Aufgabe M_G_17 (siehe Abbildung IV.23), die in das Gebiet der Vektorrechnung gehört. Zur Lösung der Aufgabe muss man Kräftediagramme interpretieren; aber auch an das Textverständnis werden recht hohe Anforderungen gestellt.

Die Zerlegung oben links wurde von 72% als richtig erkannt (δ = -1.13), die Zerlegung unten links von 89% (δ = -2.36), die Zerlegung oben rechts kreuzten 79% der Testpersonen als falsch an (δ = -1.57). Leicht schwieriger war die Zerlegung unten rechts: Diese wurde von 61% als richtig erkannt (δ = -0.59).

Abb. IV.23: Beispiel 2 einer gut gelösten Mathematikaufgabe



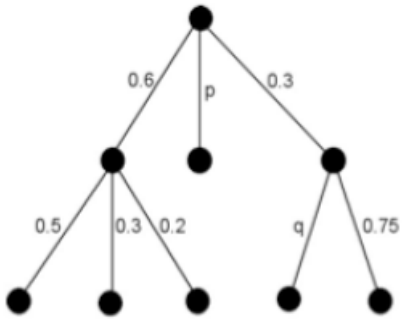
Nicht ganz unerwartet hat die Aufgabe M_S_15 (siehe Abbildung IV.24) keine grossen Probleme bereitet. Auch bei dieser Aufgabe geht es um das „Rechnen" im Gebiet der Wahrscheinlichkeitsrechnung. Die Wahrscheinlichkeit p haben 88% der Maturandinnen und Maturanden richtig bestimmt (δ = -2.23), die Wahrscheinlichkeit q 86% (δ = -2.07).

Abb. IV.24: Beispiel 3 einer gut gelösten Mathematikaufgabe

Im untenstehenden Baumdiagramm sind zwei Wahrscheinlichkeiten nicht eingetragen.

Geben Sie die Wahrscheinlichkeiten p und q an.

M_S_15



$p =$ .....

$q =$ .....

Die Aufgabe M_X_28 (siehe Abbildung IV.25) wurde ebenfalls sehr gut gelöst. Sie ist im Kompetenzraster dem Feld „Rechnen“ mit Zahlen in elementarer Algebra zugeordnet.

Abb. IV.25: Beispiel 4 einer gut gelösten Mathematikaufgabe

Welche der unten aufgeführten Zahlenpaare $(m|n)$ erfüllen die Ungleichung

$$\frac{1}{m} + \frac{1}{n} > \frac{1}{2} \ ?$$

M_X_28

$(m\|n) = (3\|4)$	☐ ja	☐ nein
$(m\|n) = (5\|3)$	☐ ja	☐ nein
$(m\|n) = (4\|4)$	☐ ja	☐ nein
$(m\|n) = (3\|3)$	☐ ja	☐ nein

Die Lösungshäufigkeiten lauten wie folgt:

- Erste Behauptung: 79% richtige Lösungen (δ = -1.49)
- Zweite Behauptung: 69% richtige Lösungen (δ = -0.97)
- Dritte Behauptung: 85% richtige Lösungen (δ = -1.98)
- Vierte Behauptung: 76% richtige Lösungen (δ = -1.36).

Mit einer Lösungshäufigkeit von über 80% folgt schliesslich noch die Aufgabe M_S_02 (siehe Abbildung IV.26). Bei dieser Aufgabe geht es um Kenntnisse von Definitionen und Sätzen im Bereich der Kombinatorik (δ = -1.63).

Abb. IV.26: Beispiel 5 einer gut gelösten Mathematikaufgabe

Ein Spielwürfel (mit sechs Seiten) wird n-mal hintereinander geworfen.

Wie viele mögliche Ausgänge hat dieser (mehr)stufige Zufallsversuch?

1 n

2 2^n

3 $6n$

4 6^n

M_S_02

Ein erstes Beispiel einer schlecht gelösten Aufgabe finder sich in Abbildung IV.27. Schwierigkeiten bereitet hat offenbar der Begriff der „Abzählbarkeit“ (δ = 2.48). Dieser Begriff aus der Mengenlehre scheint nicht geläufig zu sein: Die Lösungshäufigkeit beträgt 9%.

Abb. IV.27: Beispiel 1 einer schlecht gelösten Mathematikaufgabe

Eine Menge A ist abzählbar, wenn ...

1 wenn A endlich ist

2 wenn A unendlich viele Elemente hat

3 wenn eine Bijektion $\mathbb{N} \longrightarrow A$ existiert

4 wenn eine Bijektion $\mathbb{R} \longrightarrow A$ existiert

M_X_23

Ein ähnliches Resultat ergab Aufgabe M_X_19 (siehe Abbildung IV.28) (δ = 2.48). Es konnten ebenfalls nur 9% der Maturandinnen und Maturanden diese Aufgabe richtig lösen. Der Ausdruck „4Z“ hat vermutlich Etliche verwirrt.

Abb. IV.28: Beispiel 2 einer schlecht gelösten Mathematikaufgabe

Wir betrachten die Menge $4\mathbb{Z} = \{m \in \mathbb{Z} \mid m = 4z \text{ für ein } z \in \mathbb{Z}\}$.

Geben Sie alle ganzen Zahlen n an, die in der Menge $4\mathbb{Z}$ liegen und für die $-10 < n < 10$ gilt.

Antwort: _____

M_X_19

Die Aufgabe M_A_11 (siehe Abbildung IV.29) wurde ebenso nur von rund 10% der Testpersonen richtig beantwortet (δ = 2.38).

Abb. IV.29: Beispiel 3 einer schlecht gelösten Mathematikaufgabe

M_A_11

Für geometrische Reihen gilt die Formel:

$$\sum_{k=0}^{\infty} a^k = a^0 + a^1 + a^2 + a^3 + \cdots = \frac{1}{1-a} \text{ mit } |a| < 1$$

Wenden Sie diese Formel auf den Ausdruck

$$p\sum_{j=0}^{\infty}(1-p)^j \text{ mit } |1-p| < 1$$

an und vereinfachen Sie das Ergebnis soweit wie möglich.

Antwort: Die Summe ergibt _____

In der Abbildung IV.30 findet sich eine Aufgabe (M_A_06), deren Teile sehr unterschiedlich gut gelöst wurden. Die erste Teilaufgabe wurde von 84% der Maturandinnen und Maturanden richtig beantwortet (δ = -1.87); die zweite Teilaufgabe nur noch von 49% (δ = -0.02), die dritte Teilaufgabe von 20% (δ = 1.48) und die letzte Teilaufgabe von lediglich 11% (δ = 2.23).

Ansonsten weisen die Aufgaben aus Logik und Mengenlehre die mit Abstand grösste Streuung bezüglich Lösungshäufigkeit auf. Neben einigen sehr leichten Aufgaben gibt es hier sehr schwierige, was oft auf unbekannte Begriffe zurückzuführen ist.

Abb. IV.30: Beispiel einer Aufgabe mit unterschiedlich schwierigen Teilaufgaben

M_A_06

Gegeben sind die folgenden Funktionen

$$f(x) = x^3$$
$$g(x) = \frac{1}{x}$$
$$h(x) = -\frac{1}{x^2}$$
$$j(x) = \frac{2}{x}$$
$$k(x) = \ln\left(x^2\right)$$
$$l(x) = 3x^2$$
$$m(x) = \frac{1}{2\sqrt{x}}$$
$$n(x) = \ln(2x)$$
$$p(x) = 2^x$$

Vier Funktionen in der obigen Liste sind Ableitungen von andern Funktionen der Liste.

Vervollständigen Sie:

Die Funktion $l(x)$ ist die Ableitung der Funktion .....

Die Funktion ..... ist die Ableitung der Funktion $g(x)$.

Die Funktion $j(x)$ ist die Ableitung der Funktion .....

Die Funktion ..... ist die Ableitung der Funktion $n(x)$.

5.4 Ergebnisse in Biologie

5.4.1 Beschreibung der Gesamtergebnisse

In Tabelle IV.8 sind die Gesamtergebnisse in Biologie aufgeführt. Da an vielen Gymnasien dieses Fach bereits ein bis zwei Jahre vor den Maturitätsprüfungen abgeschlossen wird, werden die Resultate für zwei Gruppen („abgeschlossen ja“, „abgeschlossen nein“) dargestellt.

Tab. IV.8: Gesamtergebnisse Biologie

	Mittel A		S A		Max A		Mittel P		S P		Min P		Max P	
abgeschl.	ja	nein	ja	nein	ja	nein	ja	nein	ja	nein	ja	nein	ja	nein
TOT_BIO	26.8	32.5	8.71	8.84	54.2	56.3	–.22	.10	.470	.474	–1.69	–1.28	1.24	1.59
B_EVO	4.79	5.45	3.00	2.94	10.0	10.0	–.21	.09	.452	.472	–1.68	–1.68	1.21	1.35
B_STRU	3.85	4.75	2.55	2.79	10.0	10.0	–.22	.10	.460	.484	–2.07	–1.80	1.23	1.56
B_GEN	4.44	5.42	2.38	2.39	10.0	10.0	–.21	.08	.410	.428	–1.81	–1.70	1.19	1.52
B_INFO	4.16	5.28	2.46	2.48	10.0	10.0	–.24	.11	.497	.514	–1.97	–1.91	1.30	1.45
B_OEKO	4.85	5.77	2.63	2.60	10.0	10.0	–.24	.12	.533	.539	–2.22	–1.87	1.33	1.61
B_STOFF	4.56	5.78	2.53	2.56	10.0	10.0	–.25	.14	.552	.569	–2.40	–2.07	1.52	1.76

N „ja“ = 9'693; N „nein“ = 3’748 (gewichtet)

Die neu verwendeten Abkürzungen in der Tabelle verstehen sich wie folgt:

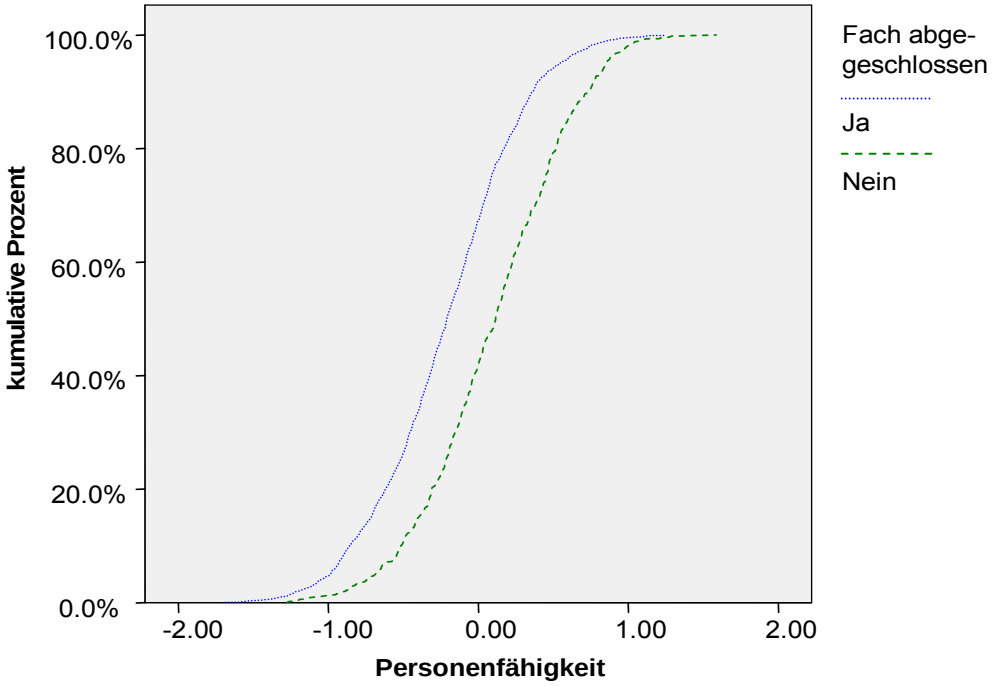
- TOT_BIO: Gesamtpunktzahl
- B_EVO: Evolution (Entwicklungsprozesse – Evolution und Zukunftsfragen)
- B_STRU: Struktur und Funktion (Bau und Funktion von Zellen, Geweben und Organen; funktionsbezogene Differenzierungen)
- B_GEN: Genetik (Grundlagen der molekularen Genetik, Anwendung moderner biologischer Erkenntnisse und Methoden)
- B_INFO: Informationsverarbeitung (Kommunikation zwischen Zellen, neuronale Informationsverarbeitung, Wahrnehmung)
- B_OEKO: Ökologie (Vernetzte Systeme – Ökologie und Nachhaltigkeit)
- B_STOFF: Stoffwechsel (Stoff- und Energiewechsel: Notwendigkeit und Wege der Energieumwandlung, Assimilation und Dissimilation im Zusammenhang mit zellulären Strukturen und dem Organismus; molekulare Steuerung von Stoffwechsel: Enzymatik)
- Mittel A: Mittelwerte nach Variante A (max. 10 Punkte pro Bereich, max. 60 Punkte total)

Im Mittel wurden die Testaufgaben in Biologie knapp zur Hälfte richtig gelöst, wobei die Gruppe mit zurückliegendem Fachabschluss noch schlechter, die andere Gruppe hingegen überdurchschnittlich abgeschlossen hat. Die Biologieaufgaben wurden von der ersten Gruppe auch im Vergleich zu Erstsprache und Mathematik eher schlechter gelöst. Das heisst, auch die geschätzten Personenfähigkeiten liegen im Durchschnitt unter einem mittleren Fähigkeitsmass. Vermutlich spielt hier der Vergessenseffekt eine Rolle, der aber auch weit höher hätte ausfallen können. Im Gesamttotal wurde die maximal mögliche Punktzahl von niemandem erreicht, wenn zum Teil auch nur knapp verfehlt. Die Streuung auf Ebene der einzelnen Personen ist wieder beachtlich. In jedem Bereich wurde auch die maximal mögliche Punktzahl erreicht. Die zusätzliche Auswertung der über die Klassen

aggregierten Werte ergibt nochmals eine erhebliche Streuung auch zwischen den Klassenmittelwerten (Minimum: B_TOTAL_B2H = 14.93; Maximum: B_TOTAL_B2H = 42.35). Wieder hat die beste Klasse im Mittel beinahe dreimal so viele Aufgaben richtig gelöst wie die schlechteste.

In Abbildung IV.31 ist die kumulative Verteilung der Ergebnisse für die Gesamtschätzung der Personenfähigkeiten grafisch dargestellt.

Abb. IV.31: Ergebnisse Personenfähigkeiten Biologie Total



5.4.2 Beschreibung der Ergebnisse ausgewählter Items

In diesem Abschnitt werden ausgewählte Items beschrieben, die unterschiedlich gut gelöst wurden. Wieder werden einerseits die effektiven, relativen Lösungshäufigkeiten und andererseits die Rasch-geschätzten objektiven Aufgabenschwierigkeiten δ angegeben.

Abb. IV.32: Beispiel 1 eines leichten Items in Biologie

(FKT_03)

Aus welchen Bausteinen sind Proteine zusammengesetzt?

- ❑ aus Nukleotiden
- ❑ aus Genen
- ❑ aus Glukosemolekülen
- ❑ aus Aminosäuren

Die Lösungswahrscheinlichkeit für das Item in Abbildung IV.32 liegt bei 0.857 (δ = -2.02). Um das Bild eines Hörsaals mit 100 Studierenden des ersten Semesters zu gebrauchen: 86 Personen würden wohl die Begriffe „Protein“ und „Aminosäure“ im Sinne von Bausteinkette und Baustein zusammenbringen. Allerdings gibt es immer noch 14 Anwesende, welche diese Assoziation wahrscheinlich nicht herstellen würden.

Abb. IV.33: Beispiel 2 eines leichten Items in Biologie

(INF_07)

Taxol aus der Eibenpflanze ist mitosehemmend und wird in der Krebstherapie eingesetzt. Wie wirkt Taxol also?

- ❑ Taxol hemmt den Zelltod der Körperzellen.
- ❑ Taxol hemmt die Bildung von Geschlechtszellen.
- ❑ Taxol hemmt die Arbeit des Immunsystems.
- ❑ Taxol hemmt die Vermehrung der Krebszellen.

Der Satz, der den Stamm der Aufgabe in Abbildung IV.33 bildet, ist unverändert dem Skript „Biodiversität: Pflanzen, Wirbeltiere, Pilze“ von Peter Linder (Universität Zürich) entnommen (S. 72). Möglicherweise hilft das Stichwort „Krebstherapie“, das zusammen mit der Sinneinheit „mitosehemmend“ erscheint, die richtige Option zu finden, weil Krebs schon in der Alltagssprache eng mit Zellvermehrung assoziiert ist. Dieser Hilfseffekt ist aber derselbe, wie er in der konkreten Lesesituation im Studium auftreten würde. Er wurde im Sinne der Gestaltung der Items im Rahmen von EVAMAR II nicht nur berücksichtigt, sondern ganz bewusst eingesetzt. Die zugrunde liegende Hypothese lautet, dass, falls die Maturandin oder der Maturand trotz der Hilfestellung den Begriff „mitosehemmend“ falsch verortet (ihn z. B. in den Zusammenhang mit Zelltod, Geschlechtszellen oder dem Immunsystem stellt), sie bzw. er beim Lesen des Skripts die Bedeutung des Adjektivs „mitosehemmend“ auch nicht verstanden hätte. Die Lösungswahrscheinlichkeit der Aufgabe liegt bei 0.809 (δ = -1.65), was bei einer 100-köpfigen Studierendenschaft bedeuten würde, dass beim Lesen des Begriffs „Mitose“ selbst im Kontext des Stichwortes „Krebstherapie“ im Schnitt 19 Studentinnen und Studenten diesen nicht mit „Zellvermehrung“ assoziieren würden. Dieses Item diskriminiert trotz seines geringen Schwierigkeitsgrades in beiden Sprachen recht gut (Diskrimination > 0.25). Auf den gesamten Biologietest bezogen sind es also vor allem schwache Maturandinnen und Maturanden, die hier falsche Optionen gewählt haben.

Abb. IV.34: Beispiel 3 eines leichten Items in Biologie

(OEK_03)

Ein Parasit schadet einem anderen Lebewesen, zieht aber selber aus diesem Zusammenleben einen Nutzen. Welche Präzisierung kann noch hinzugefügt werden?

- ❑ Er lebt auf oder in dem anderen Lebewesen.
- ❑ Er lebt auf dem anderen Lebewesen.
- ❑ Er lebt in dem anderen Lebewesen.
- ❑ Er kommt mit dem anderen Lebewesen nicht direkt in Berührung.

Von 100 Maturandinnen und Maturanden konnten bei der Aufgabe in Abbildung IV.34 etwa 77 das Kreuz an der richtigen Stelle setzen, sind sich also bewusst, dass Parasiten durchaus auch in ihren Wirten leben können (so genannte Endoparasiten) (δ = -1.43). Die Studentinnen und Studenten könnten bei einem Endoparasiten z. B. an den berühmten Bandwurm gedacht haben. Andere Endoparasiten, die nicht zum Alltagswissen gezählt werden dürfen, sind zum Beispiel die einzelligen Erreger der Malaria (Plasmodium sp.), die rote Blutkörperchen von Wirbeltieren bewohnen, oder der Erreger der Schlafkrankheit, ein Geisseltierchen, das im Blut des Menschen lebt. Es kann hier nicht entschieden werden, ob der hohe Lösungsgrad dieser Aufgabe auf dem hilfreichen Hinweis aus dem Alltagswissen (Bandwurm) beruht oder auf der Beherrschung tatsächlich weitergehenden Biologiewissens (wie es das Wissen über Malaria-Erreger und die Schlafkrankheit darstellt).

Abb. IV.35: Beispiel 4 eines leichten Items in Biologie

(STW_09)

Bei welchem Vorgang spielen Chloroplasten eine zentrale Rolle?

- ❑ bei der Zellteilung
- ❑ bei der Zellatmung
- ❑ bei der Produktion von ATP
- ❑ bei der Photosynthese

Auf dem in der Aufgabe in Abbildung IV.35 abgefragten Wissensinhalt sollte eine Dozentin oder ein Dozent an der Universität eigentlich mit Sicherheit aufbauen können. Tatsächlich assoziierten 75.5% der Maturandinnen und Maturanden mit dem Begriff „Chloroplasten" den Vorgang der Photosynthese. Selbstverständlich hat auch die Zellatmung (Option 2) eine gewisse Anziehungskraft, da Maturandinnen und Maturanden Vorgänge der Assimilation und Dissimilation sehr oft verwechseln (δ = -1.32).

Abb. IV.36: Beispiel 1 eines mittelschwierigen Items in Biologie

(STW_07)

„Alle Tiere müssen die gleichen grundsätzlichen Lebensaufgaben erfüllen, u. a. auch die Kontrolle von Wasserhaushalt und Salzkonzentration (Osmoregulation)".

Welche beiden Aussagen zum Vorgang der Osmose sind korrekt?

I. Salzteilchen wandern durch die semipermeable Membran.
II. Wassermoleküle wandern durch die semipermeable Membran.
III. Die semipermeable Membran entspricht bei Lebewesen einer Zellmembran.
IV. Die Durchquerung der semipermeablen Membran verbraucht Energie (ATP).

- ❑ I und IV
- ❑ I und III
- ❑ II und III
- ❑ II und IV

Das Item in Abbildung IV.36 lösten gerade 51.7% der befragten Maturandinnen und Maturanden richtig (Antwort 3) (δ = -0.16). Da korrekte Assoziationen zu einer Textstelle, welche die Vorsilbe „osmo-" enthält, für das Verständnis fast aller physiologischer Prozesse absolut grundlegend sind, erstaunt es hier, dass der Prozentsatz nicht höher liegt. Zwei Gründe sind naheliegend:

c) Die Vermittlung des Konzepts der Osmose gehört zu den heikelsten und aufwändigsten Unterfangen im Biologieunterricht. Das Konzept ist eine Modellvorstellung über das Verhalten von unsichtbaren kleinsten Teilchen und stellt hohe Anforderungen an die Vorstellungskraft und an das Abstraktionsvermögen.

d) Die Vermittlung der Konzepte der Diffusion und Osmose findet (ausgerechnet wegen ihrer grundlegenden Bedeutung für die gesamte Biologie) in der Regel zu Beginn des gymnasialen Biologieunterrichts statt und scheint dann sowohl für die Maturandinnen und Maturanden als auch für manche Lehrpersonen „abgehakt" zu sein. Gerade der Umgang mit sinnlich nicht wahrnehmbaren Teilchen sowie mit Abs-

traktion wird im Verlauf der gymnasialen Ausbildung aber erst entfaltet. Oft fehlt dann bis zur Matura ein „unerbittliches“ Wiederaufgreifen der Prozesse Diffusion und Osmose, um sie einerseits wieder in Erinnerung zu rufen, und um andererseits Misskonzepte in immer neuen Aufgabenkontexten sicher zu entlarven und zu überwinden.

Abb. IV.37: Beispiel 2 eines mittelschwierigen Items in Biologie

(GEN_07)

Was ist eine Zygote?

- ❑ Eine Zygote ist ein Embryo noch vor der Einnistung in die Gebärmutter.
- ❑ Eine Zygote ist eine Zelle entstanden durch die Verschmelzung zweier Keimzellen.
- ❑ Eine Zygote ist ein Embryo im Zweizellstadium.
- ❑ Eine Zygote ist eine Eizelle, die nicht befruchtet wurde.

Auch das Item in Abbildung IV.37 konnte etwa die Hälfte der Maturandinnen und Maturanden (51.7%) am Ende der gymnasialen Ausbildung korrekt beantworten (δ = -0.20).

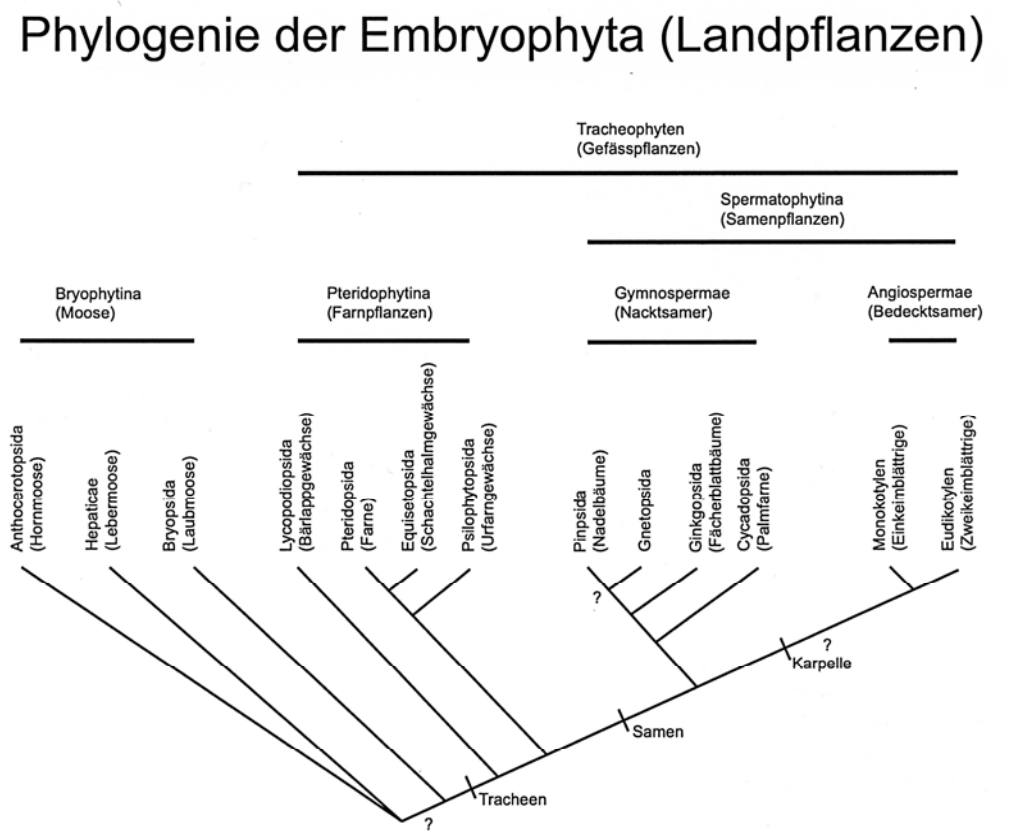
Bei diesem Item stellt sich die Frage, wie gross der Anteil an Lehrpersonen ist, die aus ihrer Sicht auf „unnötige“ Fremdworte im Sinne einer Vermeidung nicht erforderlicher kognitiver Belastung („cognitive load“) konsequent verzichten. Verwendet die Lehrperson anstelle von „Zygote“ stets den deutschsprachigen Ausdruck „befruchtete Eizelle“, dann leidet das Verständnis wichtiger Basiskonzepte (wie zum Beispiel der zweigeschlechtlichen sexuellen Fortpflanzung mit zwingender Gametenverschmelzung) sicherlich nicht. Im Sinne der Studierfähigkeit erscheint die Vermeidung „unnötiger“ kognitiver Belastung aber eher wie ein Aufschieben derselben, sofern hier angenommen wird, dass bei Beginn des Studiums der Begriff „Zygote“ auf jeden Fall bekannt sein soll. Die Notwendigkeit, viele Fachbegriffe beim Lesen eines Skripts auf Hochschulniveau vermehrt nachschlagen zu müssen, kann sich zu Studiumsbeginn unter Umständen schnell akkumulieren, was dann häufig mit (zuweilen grossen und bedrohlichen) Lernschwierigkeiten einhergeht.

Das Item in Abbildung IV.38 erfordert das Anwenden der Fertigkeit des Lesens von Stammbäumen (δ = -0.04). 48.1% der Probandinnen und Probanden zeigten mit ihrer Antwort, dass sie die spezifische phylogenetische Symbolsprache verstanden haben, die an der Universität vorausgesetzt wird (Antwort 2). Wieder ist zu vergegenwärtigen, dass in einem Hörsaal von 100 Studentinnen und Studenten deren 52 nicht in der Lage sind, auf korrekte Art und Weise Informationen aus einer phylogenetischen Abbildung zu ziehen.

Abb. IV.38: Beispiel 3 eines mittelschwierigen Items in Biologie

Item 3 (EVO_07)

Welche Aussage kann man aus der nachfolgenden Graphik herauslesen?



- ❑ Nadelbäume produzieren Samen, Einkeimblättrige aber nicht.
- ❑ Nadelbäume und Einkeimblättrige produzieren Samen.
- ❑ Der gemeinsame Vorfahre der Farne und der Nadelbäume produzierte bereits Samen.
- ❑ Alle heutigen Landpflanzen produzieren Tracheen, Samen sowie Karpelle.

Abb. IV.39: Beispiel 4 eines mittelschwierigen Items in Biologie

(OEK_20)

Prochlorococcus marinus **bildet einzelne kugelige Zellen (0.6–0.8 µm Durchmesser).**
Die Algenart ist im Plankton der Meere weit verbreitet. Was bedeutet das?

- ❑ Die Algenart bildet frei im Wasser treibende und schwebende Zellen.
- ❑ Die Algenart lebt nur im Süsswasser.
- ❑ Die Algenart lebt in Küstennähe auf dem Meeresgrund.
- ❑ Die Algenart bildet neue Nährstoffe aus toten Organismen.

Die Textstelle aus einem Erstsemestrigen-Skript in der Aufgabe in Abbildung IV.39 enthält den Fachbegriff „Plankton“. Mehr als die Hälfte der Maturandinnen und Maturanden (52.8%) konnte mit dem zweiten Satz etwas anfangen (δ = -0.19). Für die restlichen 47.2% ruft der Satz nicht korrekte Assoziationen zum Lebensraum (Option 2: Süsswasser, Option 3: Meeresgrund) und der ökologischen Rolle (Option 4: Destruent statt Produzent) hervor.

Die folgenden Aufgaben wurden alle eher schlecht gelöst.

Abb. IV.40: Beispiel 1 eines schwierigen Items in Biologie

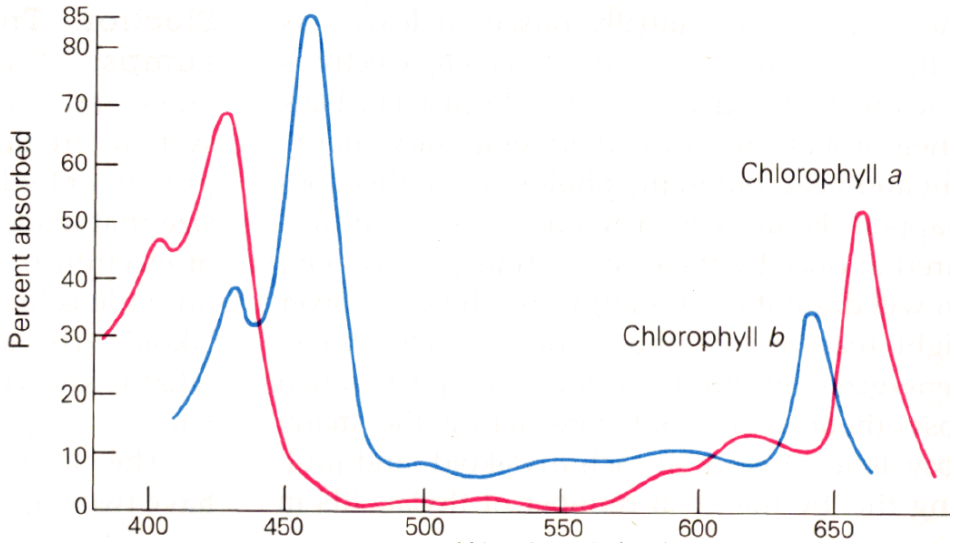
(STW_21)

Im untenstehenden Diagramm liegen zwei Kurven in einem Koordinatensystem. Die x-Achse trägt eine typische Skala und Grösseneinheit und ebenso trägt die y-Achse eine typische Skala und Grösseneinheit.

Wie heisst die Eigenschaft von Chlorophyll a bzw. Chlorophyll b, welche immer auf diese Weise dargestellt wird?

(Antwort: nur genau 1 Stichwort, kein Satz)

_____



Bei der Aufgabe in Abbildung IV.40 war im Gegensatz zu den übrigen hier beschriebenen Beispielen ein Fachausdruck gefragt, den die Probandin beziehungsweise der Proband aktiv erinnern musste.

Die gesuchten Begriffe waren in deutscher Sprache „Absorptionsspektrum“ und in französischer Sprache „spectre d’absorption“, „spectre d’absorption de l’énergie lumineuse“ oder „capacité d’absorption du spectre lumineux“. Nur 10.7% der befragten Personen nannten einen dieser Begriffe (δ = 1.78). Die Aufgabenstellung, einen biologischen Fachbegriff zu suchen, der Auskunft über das Absorptionsvermögen nicht nur bei einer bestimmten Wellenlänge, sondern aufgrund seiner Definition zwingend über einen ganzen Bereich des Spektrums geben muss, erwies sich für viele als zu anspruchsvoll. Neben den vielen leer gelassenen Antwortzeilen (44% in

deutscher und 64% in französischer Sprache) zeigt sich dieses Defizit bzw. eine gewisse Hilflosigkeit auch in der semantischen Distanz vieler falscher Antworten vom korrekten Begriff. Hier seien einige Beispiele in deutscher Sprache aufgeführt: Aufnahmefähigkeit, Durchlässigkeit des Lichts, Energiegewinnung, Farbe, Lichtintensität, Photosynthese/Fotosynthese und Wirkungsspektrum. In französischer Sprache wurden irrtümlicherweise genannt: couleur, photosynthèse, luminosité und pigmentation.

Abb. IV.41: Beispiel 2 eines schwierigen Items in Biologie

(INF_01)

Die Netzhaut hat auf dem Corpus geniculatum laterale eine neuronale Karte. Was bedeutet der Begriff der „neuronalen Karte"?

- ❑ Die Aktivität der Sinneszellen in der Netzhaut wird auf einem zweidimensionalen Areal im Corpus geniculatum laterale abgebildet.
- ❑ Die Aktivität der Sinneszellen in der Netzhaut wird im Corpus geniculatum laterale zeitweise gespeichert.
- ❑ Die Aktivität jeder einzelnen Sinneszelle der Netzhaut wird vom Corpus geniculatum laterale gezielt gesteuert.
- ❑ Die Aktivität jeder einzelnen Zelle des Corpus geniculatum laterale wird von den Sinneszellen der Netzhaut gezielt gesteuert.

Beim in Abbildung IV.41 dargestellten Item liegt der Prozentsatz der richtigen Antworten sowohl in deutscher Sprache (22.5%) als auch in französischer Sprache (12.8%) unter der rechnerischen Wahrscheinlichkeit von 25%, die richtige Antwort (Option 1) durch reines Raten zu treffen (δ = 1.41). Die Distraktoren scheinen insgesamt attraktiver zu sein als die korrekte Option. Beim Betrachten der Diskriminations-Werte der einzelnen Distraktoren fällt auf, dass die Werte bei den Distraktoren in beiden Sprachen allesamt nahe bei null liegen. Das heisst, dass sowohl die auf den gesamten Biologietest bezogen schwachen wie auch die starken Maturandinnen und Maturanden dieselben falschen Distraktoren als attraktiv beziehungsweise unattraktiv einstufen. Dies könnte dahingehend interpretiert werden, dass bei diesem Item schwache und starke Maturandinnen und Maturanden auf ähnliche Weise raten, wenn sie die Antwort nicht „zufällig" wissen, dass also auch starke Maturandinnen und Maturanden den Distraktoren nicht anmerkten, dass sie falsch sind. Auch liessen auffällig viele Probandinnen und Probanden diese Aufgabe aus, verzichteten also darauf, ein Kreuz zu setzen (D: 16.3%, F: 15.8%). Dies ist ein zusätzlicher Hinweis darauf, dass es sich hier eindeutig um eine sehr schwierige Aufgabe handelt. Der Begriff der „neuronalen Karte" wird von einem überwiegenden Teil der Maturandinnen und Maturanden nicht als Eingangswissen mitgebracht.

Abb. IV.42: Beispiel 3 eines schwierigen Items in Biologie

(FKT_13)

Die Fruchtkörper gewisser Protozoen haben morphologische Ähnlichkeiten mit Pilzfruchtkörpern. Von welcher Art der Ähnlichkeit ist die Rede?

- ❑ von der Ähnlichkeit der Vermehrungsrate
- ❑ von der Ähnlichkeit der Evolutionsgeschichte
- ❑ von der Ähnlichkeit des Stoffwechsels in den Fruchtkörperzellen
- ❑ von der Ähnlichkeit des Aussehens

Sehr ähnlich liegen die Verhältnisse beim Item in Abbildung IV.42. Die Prozentzahlen für die korrekte Option liegen in beiden Sprachen unter der Ratewahrscheinlichkeit von 25% (D: 24.2%, F: 17.7%) (δ = 1.23). Die Diskriminations-Werte der falschen Optionen befinden sich ebenfalls alle sehr nahe bei null. Auch hier scheinen die Maturandinnen und Maturanden „verloren" zu sein, wenn sie zum Raten gezwungen sind, das heisst, auch biologisch versierte Maturandinnen und Maturanden wählen genauso wie schwache Maturandinnen und Maturanden überzufällig die falsche Option 3 (D: 34.9%, F: 43.0%). Dies könnte damit zusammenhängen, dass ausschliesslich Option 3 den Fachbegriff „Fruchtkörper" aus dem Aufgabenstamm noch einmal wiederholt. Die Rateheuristik der Stamm-Optionen-Wortwiederholung scheint unabhängig vom allgemeinen biologischen Können der Probandin beziehungsweise des Probanden zum Zuge gekommen zu sein. Eine Dozentin oder ein Dozent sollte also in einer Erstsemestrigenvorlesung den Begriff „morphologisch" als „die äusserlich sichtbare Form und Struktur betreffend" nicht unkommentiert anwenden, da im Hörsaal die Assoziationen dazu in den meisten Fällen nicht korrekt ausfallen.

Abb. IV.43: Beispiel 4 eines schwierigen Items in Biologie

(FKT_01)

Welche Merkmale gehören zu der <u>anatomischen</u> Beschreibung des Verdauungsapparats einer Kuh?

I. das Gesamtgewicht des Dünndarmes
II. die Gesamtlänge der Speiseröhre
III. der pH-Wert im Magen
IV. der Gewichtsanteil der im Darmtrakt lebenden Bakterien
V. der Methanausstoss pro Stunde
VI. die dickste Stelle der Dünndarmschleimhaut

- ❑ I, II, VI
- ❑ I, III, VI
- ❑ I, II, IV, V
- ❑ I, III, IV, V

Das Item in Abbildung IV.43 tangiert eine bio-ontologisch auf einer sehr hohen Hierarchie-Ebene angesiedelte Unterscheidung (δ = 1.20). Die anatomische Betrachtungsweise mit Fokus auf äussere oder innere Form und Struktur wird in der Biologie abgegrenzt von einer physiologischen Betrachtungsweise mit Fokus auf die Vorgänge, die basierend auf diesen äusseren und inneren Formen und Strukturen ablaufen. Merkmal III ist also kein anatomisches Merkmal, weil es sich beim pH-Wert weder um eine Struktur noch um eine Form handelt. Ebenso verhält es sich mit einem Gewichtsanteil (Merkmal IV) oder dem Ausstossvolumen eines Stoffes pro Stunde (Merkmal V). Damit verbleiben die Merkmale I, II und VI als anatomische Merkmale. Option 1 ist somit korrekt. Es fällt auf, dass die Maturandinnen und Maturanden alle vier Optionen sehr ähnlich häufig gewählt haben. Besonders in der französischen Fassung des Items sind die Häufigkeiten nahezu gleich verteilt (Option 1: 22%, Option 2: 20%; Option 3: 24%, Option 4: 22%). Aber auch in der deutschsprachigen Version zeigt sich ein auffällig ausgeglichenes Bild (Option 1: 22%, Option 2: 21%, Option 3: 18%, Option 4: 30%). Die restlichen 9% beziehungsweise 12% für die französischsprachige Fassung entfallen auf diejenigen, die das Item übersprungen haben. Die korrekte Option vereint als deutliches Zeichen eines offenbar sehr schwierigen Items wiederum weniger als 25% der Antworten auf sich. Auch hier kann im Anschluss an die gymnasiale Ausbildung in Vorlesungen zu Studienbeginn nicht davon ausgegangen werden, dass mit der Nennung des Begriffs „anatomisch" automatisch die Gedanken in die richtige Richtung gehen.

5.5 Zum Vergleich: Erzielte Maturanoten

Die Auswertung der Testergebnisse offenbart in allen Bereichen eine erstaunlich breite Streuung, vor allem angesichts der Tatsache, dass die Maturandinnen und Maturanden kurz vor der Verleihung der für alle Studienfächer geltenden, universalen Qualifikation „Hochschulreife“ standen. Es kann deshalb davon ausgegangen werden, dass nicht alle Maturandinnen und Maturanden in allen drei getesteten Fachbereichen über Kompetenzen verfügen, die den universitären Anforderungen aller Studienfächer entsprechen. Deshalb soll zum Vergleich auch ein Blick auf die Mittelwerte und die Streuungen effektiv vergebener Maturanoten in den drei Testfächern geworfen werden (siehe Tabelle IV.9).

Tab. IV.9: ausgewählte Maturanoten

Prüfungen	N (gewichtet)	Min	Max	Mittel	S
Erstsprache schriftlich	13’394	2.00	6.00	4.41	.739
Erstsprache Gesamtnote	13’390	2.50	6.00	4.63	.542
Mathematik schriftlich	13’390	1.00	6.00	4.03	1.048
Mathematik Gesamtnote	13’381	1.50	6.00	4.34	.826
Biologie Erfahrungsnote	11’112	2.50	6.00	4.68	.541
Naturwissenschaften Gesamtnote	13’369	2.50	6.00	4.60	.566

Ausser in den schriftlichen Mathematikprüfungen wurde der minimal mögliche Wert (Note 1) nicht vergeben. Die relative Streuung fällt insgesamt kleiner aus als bei den Tests. Es ist zudem bemerkenswert, dass die Streuung in Mathematik erheblich grösser ist als in den beiden anderen Fächern. Ebenso, dass in den schriftlichen Mathematikprüfungen die Durchschnittsnote bei 4.0, also gerade noch bei „genügend“ liegt. Eine Auswertung der über die Klassen aggregierten Zahlen ergibt folgende Extremwerte der Klassendurchschnitte: Minima: Erstsprache schriftlich 3.6, Erstsprache Gesamtnote 3.97, Mathematik schriftlich 2.88, Mathematik Gesamtnote 3.73, Biologie Erfahrungsnote 3.85, Naturwissenschaften Gesamtnote 4.0; Maxima: Erstsprache schriftlich 5.11, Erstsprache Gesamtnote 5.17, Mathematik schriftlich 5.48, Mathematik Gesamtnote 5.44, Biologie Erfahrungsnote 5.28, Naturwissenschaften Gesamtnote 5.31.

Bei Notenbewertungen steht fest, dass Noten unter 4.0 als ungenügende Qualifikation gelten. Demgemäss können aus einer weiteren Auswertung der Notendaten die nachfolgend aufgeführten Prozentanteile von Maturandinnen und Maturanden angegeben werden, die in einzelnen Bereichen von den Schulen beziehungsweise den entsprechenden Lehrpersonen selbst als ungenügend qualifiziert wurden (Note 3.9 oder tiefer).

- Schriftliche Erstsprachprüfung: 19.6%
- Gesamt-Maturanote in Erstsprache: 4.7%
- Schriftliche Mathematikprüfung: 41.4%
- Gesamt-Maturanote in Mathematik: 24.4%
- Biologie-Erfahrungsnote: 5.6%
- Gesamt-Maturanote in Naturwissenschaften: 5.6%.

In der schriftlichen Maturaprüfung in Mathematik wurden also 41.4% der Schweizer Maturandinnen und Maturanden als ungenügend bewertet. 19.6% erzielten in der schriftlichen Erstsprachprüfung (überwiegend in Form eines Aufsatzes) ungenügende Noten.

Die Zahlen steigen beim Übergang in den gerade noch genügenden Bewertungsbereich (Note 4) sprunghaft an. Dies sei im Folgenden bei den Noten mit Zehntelabstufungen beim Übergang von der Note 3.9 zur Note 4.0 gezeigt:

- Schriftliche Erstspracheprüfung: Sprung von 19.6% auf 37.9%
- Schriftliche Mathematikprüfung: Sprung von 41.4% auf 56.7%
- Biologie-Erfahrungsnote: Sprung von 5.6% auf 16.7%.

Vermutlich neigt man bei dieser Grenze zum Aufrunden, so dass die objektive Zahl der ungenügenden Qualifikationen eher noch höher ausfallen dürfte.

Die oben aufgrund der Leistungstestergebnisse von EVAMAR II angebrachte Vermutung partiell ungenügender Kompetenzen bestätigt deshalb lediglich eine – zumindest teilweise – auch von den Examinatorinnen und Examinatoren ausgeübte Bewertungspraxis. Einzig die genauen Zahlen fallen anders aus. Bei der Notengebung dürfte zudem eine gewisse Anpassung an die Leistungsfähigkeit von Klassen eine Rolle spielen.

5.6 Ergebnisse „Überfachlicher Fähigkeitstest" (UFT)

In Tabelle IV.10 sind die Gesamtergebnisse für die vier Bereiche des – wie oben erwähnt, nur in der Deutschschweiz eingesetzten – UFTs aufgeführt. Die Werte für die einzelnen Kompetenzbereiche ergeben sich aus der Addition der Ergebnisse in den einzelnen Aufgaben. Es ist zu beachten, dass nicht alle Maturandinnen und Maturanden alle Kompetenzbereiche bearbeitet haben. Die Verteilung der Testhefte war aber gleichmässig.

Tab. IV.10: Gesamtergebnisse Überfachlicher Fähigkeitstest (UFT)

Bereiche	Mittel A	S A	Max A	Mittel P	S P	Min P	Max P
U_QUANT	4.57	2.496	10.00	.06	.839	–2.93	2.55
U_DIAGR	5.28	2.270	10.00	.08	.779	–2.37	2.33
U_TEXT	5.75	2.512	12.00	.09	.717	–2.39	2.32
U_NAT	4.58	2.191	10.00	.04	.712	–2.39	2.41

N = 9'281 (gewichtet)

Die neu verwendeten Abkürzungen in der Tabelle bedeuten Folgendes:

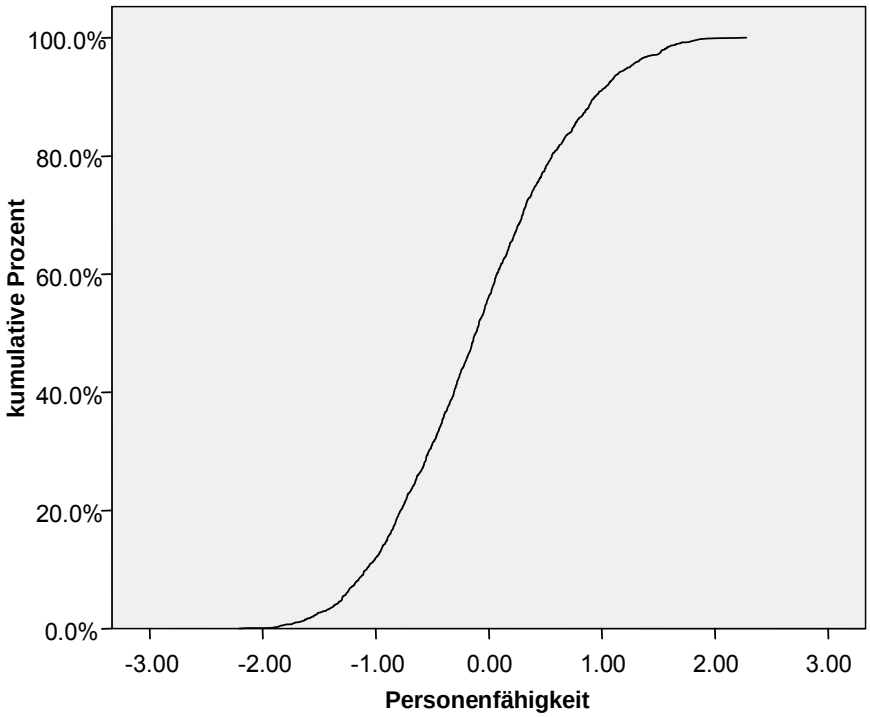
- U_QUANT: Quantitative und formale Probleme
- U_DIAGR: Diagramme und Tabellen
- U_TEXT: Textverständnis
- U_NAT: Naturwissenschaftliches Grundverständnis
- Mittel A: Erreichte Maximalpunktzahl Variante A (max. 10 Punkte pro Bereich, 12 Punkte bei U_TEXT)

Im Mittel wurden die Aufgaben des UFT in den Bereichen „Quantitative und formale Probleme", „Textverständnis" und „Naturwissenschaftliches Grundverständnis" zu etwas weniger als der Hälfte richtig gelöst. Einzig im Bereich „Diagramme und Tabellen" liegen die Werte etwas über der Hälfte. Insgesamt befinden sich die Werte leicht unter den Ergebnissen des vergleichbaren Eignungstests für das Medizinstudium (EMS). Der tiefere Wert lässt sich damit erklären, dass bei Weitem nicht alle Maturandinnen und Maturanden den EMS ablegen und die entsprechende freiwillige Vorselektion bereits zu besseren Ergebnissen führt. Diese Vergleichswerte deuten auch darauf hin, dass sich die Probandinnen und Probanden offenbar angestrengt haben. Die Ergebnisse im UFT sind somit auch als ein wichtiger Indikator für eine überwiegend seriöse Bearbeitung aller Tests zu deuten. Die geschätzten Personenfähigkeiten liegen im Durchschnitt in allen Kompetenzbereichen auf einem Niveau, auf

dem die Maturandinnen und Maturanden Aufgaben lösen können, deren Anforderungen leicht über einem mittleren Schwierigkeitsgrad liegen. Auffällig ist wieder die breite Streuung. In allen Bereichen wurde von einzelnen Maturandinnen und Maturanden auch das Punktemaximum erreicht. Eine zusätzliche Auswertung der über die Klassen aggregierten Werte ergibt die im Vergleich zu den anderen Tests grössten Differenzen zwischen den Mittelwerten der jeweils besten und schlechtesten Klasse (Minima: U_QUANT = 1.60, U_DIAGR = 2.50, U_TEXT = 2.44, U_NAT = 2.25; Maxima: U_QUANT = 8.30, U_DIAGR = 8.40, U_TEXT = 9.56, U_NAT = 7.67).

In Abbildung IV.44 ist die kumulative Verteilung der Ergebnisse für die Schätzung der Personenfähigkeiten im Gesamttest grafisch dargestellt.

Abb. IV.44: Ergebnisse Personenfähigkeiten UFT Total



5.7 Ergebnisse Zusatzfragebogen

Es folgen einige Ergebnisse des Zusatzfragebogens, die sich auf die Einschätzung der Tests aus Sicht der Maturandinnen und Maturanden und auf die überfachlichen Kompetenzen beziehen.

5.7.1 Einschätzung der Tests

Die Mittelwertsergebnisse zur Frage an die Maturandinnen und Maturanden, wie interessant sie die verschiedenen Tests fanden (von 1 = nicht interessant bis 4 = sehr interessant) zeigen, dass die Tests im Durchschnitt insgesamt als eher nicht interessant empfunden wurden. Dabei treten bei der Gesamtstichprobe geringe Unterschiede bezüglich der Einschätzung der drei Tests auf (Mathematik: M = 2.15, SD = .92, Biologie: M = 2.13, SD = .95 und Erstsprache: M = 2.16, SD = .87). Beim Vergleich der beiden Sprachregionen zeigen sich signifikante Unterschiede in allen drei Fächern: Während die Deutschschweizer Maturandinnen und Maturanden den Erstsprachetest (M = 2.22, SD = .88) sowie den Biologietest (M = 2.21, SD = .96) signifikant als interessanter einstuften als ihre französischsprachigen Pendants (mit M_{Erst} = 2.03, SD_{Erst} = .83, und M_{Bio} = 1.97, SD_{Bio} = .90), stiess der Mathematiktest in der Romandie auf mehr Interesse (M = 2.26, SD = .91) als in der Deutschschweiz (M = 2.10, SD = .93). Signifikant als am interessantesten wird der Test der überfachlichen Fähigkeiten bewertet,

der ja nur bei den deutschsprachigen Maturandinnen und Maturanden erhoben wurde; der Mittelwert beträgt hier 2.39 (SD = .93) und befindet sich damit am nächsten am theoretischen Mittelwert von 2.5.

Die insgesamt eher distanzierte Haltung gegenüber den Tests wirkte sich jedoch offensichtlich nicht auf die Anstrengungsbereitschaft der Maturandinnen und Maturanden beim Ausfüllen der Testfragen aus. Der Mittelwert von 2.87 (SD = .67) einer aus drei Items gebildeten Skala, welche die Sorgfalt, die Konzentration und das Engagement der Maturandinnen und Maturanden bei der Bearbeitung der Testaufgaben thematisiert, verdeutlicht dies. Er liegt über dem theoretischen Mittelwert von 2.5 und signalisiert damit, dass die Maturandinnen und Maturanden insgesamt die Tests eher motiviert ausgefüllt haben. Dieser Befund gilt für beide Sprachregionen, Deutschschweizer und Romands unterscheiden sich auch nicht in ihrer Anstrengungsbereitschaft.

5.7.2 Wahrnehmung überfachlicher Kompetenzen

Wie bereits in Abschnitt 2.9 beschrieben, wurden folgende überfachliche Kompetenzen, die – auch gemäss Befragung der Dozierenden in Teilprojekt A4 – für den Studienalltag als wichtig erachtet werden, in die Befragung aufgenommen: Zeitplanung, zeiteffiziente Prüfungsvorbereitung, Zeiteffizienz, Konzentrationsfähigkeit, vernetztes Denken, Diskussionsfähigkeit, Perspektivenübernahme. Die Maturandinnen und Maturanden sollten auf einer Skala von 1 (= sehr selten) bis 5 (= sehr oft) beziehungsweise für die Diskussionsfähigkeit und die Perspektivenübernahme von 1 (= trifft nicht zu) bis 4 (= trifft völlig zu) angeben, inwiefern Aussagen zu den genannten überfachlichen Kompetenzen auf sie selbst zutreffen. In Tabelle IV.11 finden sich die Mittelwerte und Standardabweichungen der verschiedenen Skalen, für die sich befriedigende Werte bei der Reliabilitätsprüfung ergaben.

Tab. IV.11: Mittelwerte der Skalen der überfachlichen Kompetenzen

Kompetenzbereiche	Min	Max	Mittel	S
Zeitplanung	1.00	5.00	2.64	.961
Effiziente Prüfungsvorbereitung	1.00	5.00	3.66	.666
Zeiteffizienz	1.00	5.00	3.76	.646
Konzentrationsfähigkeit	1.00	5.00	2.86	.700
Vernetztes Denken	1.00	5.00	3.29	.777
Diskussionsfähigkeit	1.00	4.00	3.28	.513
Perspektivenübernahme	1.00	4.00	2.99	.578

N max. = 13'358 (gewichtet)

Wie sich zeigt, sind die selbst berichteten überfachlichen Kompetenzen unterschiedlich stark ausgeprägt. Während die Maturandinnen und Maturanden ihrer Einschätzung zufolge „selten bis manchmal“ aktive Zeitplanung betreiben (M = 2.64) und auch angeben, sich nur „selten bis manchmal“ auf eine Sache konzentrieren zu können (M = 2.86), haben sie offensichtlich wenig Schwierigkeiten, sich effizient auf Prüfungen vorzubereiten (M = 3.66) oder ihre Zeit effizient zu nutzen (M = 3.76). Bezüglich ihres Verhaltens bei Teamarbeiten schätzen sich die Maturandinnen und Maturanden als kompetent ein: die Fragen danach, ob sie in Gruppenarbeiten die Perspektive anderer Gruppenmitglieder übernehmen, beantworten sie im Durchschnitt mit „trifft eher zu“, und noch ausgeprägter halten sie sich (mit einem Mittelwert von 3.28) für sehr diskussionsfähig.

Die Ergebnisse der Selbsteinschätzung von Komponenten der Ausdauer sind aus Tabelle IV.12 ersichtlich. Für diese Dimension der überfachlichen Kompetenzen liess sich keine reliable Skala bilden, so dass die Ergebnisse auf der Ebene der Einzel-Items berichtet werden. Es ist ersichtlich, dass Maturandinnen und Maturanden im

Mittel „oft“ ein ausdauerndes Lernverhalten zeigen (Items 1–3), dabei aber nur „selten“ glauben, mehr Zeit als die Mitschülerinnen und -schüler aufzuwenden (Item 4).

Tab. IV.12: Mittelwerte der Fragen zur Ausdauer (1 = sehr selten; 5 = sehr oft)

	Mittel	S
Auch wenn der Lernstoff sehr schwierig oder komplex ist, gebe ich nicht auf.	3.70	0.839
Ich lerne auch spätabends und am Wochenende, wenn es sein muss.	3.85	1.086
Ich lerne so lange, bis ich mir sicher bin, den Stoff verstanden zu haben.	3.63	0.836
Ich nehme mir mehr Zeit zum Lernen als die meisten meiner Mitschüler/innen.	2.27	1.087

N max. = 13'350 (gewichtet)

Insgesamt nehmen sich die Maturandinnen und Maturanden der Gesamtstichprobe beim Lernen als ausdauernd wahr, und sie schätzen ihre Mitschülerinnen und Mitschüler ähnlich ein. Denn beim erfragten Vergleich mit diesen geben sie an, nicht mehr Zeit zum Lernen aufzuwenden als die meisten Mitschülerinnen und Mitschüler (M = 2.27). Dies erscheint konsequent, denn offensichtlich ist es unter den Maturandinnen und Maturanden durchaus üblich, auch spätabends oder am Wochenende zu lernen, wenn es sein muss (M = 3.85). Des Weiteren stimmen die Schülerinnen und Schüler überwiegend den Aussagen zu, dass sie nicht aufgeben, auch wenn der Lernstoff sehr schwierig oder komplex ist (M = 3.70), oder sie so lange lernen, bis sie sich sicher sind, den Stoff verstanden zu haben (M = 3.63). Alles in allem stellen sich die Maturandinnen und Maturanden damit ein gutes Zeugnis in Bezug auf ihre Ausdauer beim Lernen aus.

5.7.3 Fachinteressen und Studienpropädeutik (nur Deutschschweiz)

Als nächstes werden einige Ergebnisse zu Fragen berichtet, die (aus Mangel an Erhebungszeit in der französischen Schweiz und im Tessin) nur in der Deutschschweiz erhoben wurden. Berichtet werden sollen Ergebnisse von Fragen zum Fachinteresse sowie von Fragen zur Studienvorbereitung durch Schule und Unterricht.

Das Fachinteresse wurde über insgesamt 13 Items erhoben, die – Faktorenanalysen zufolge – für alle drei Schulfächer Mathematik, Deutsch und Biologie verschiedene Dimensionen der Motivation im Unterricht abdecken: intrinsische Motivation, instrumentelle Motivation sowie leistungsbezogene Orientierung. Alle 13 Items lassen sich aber auch zu einer Skala „Interesse am Fach“ zusammenfassen (interne Konsistenzen: zwischen .90 und .91). Für die so gebildeten Interessen-Skalen ergeben sich die in Tabelle IV.13 dargestellten Mittelwerte, wobei hohe Werte für ein hohes Interesse stehen.

Tab. IV.13: Interesse am Unterrichtsfach

	Minimum	Maximum	Mittelwert	Standardabweichung
Interesse an Mathematik	1.00	4.00	2.04	.600
Interesse an Biologie	1.00	4.00	2.41	.606
Interesse an Deutsch	1.00	4.00	2.27	.633

N max. = 9’293 (gewichtet)

Insgesamt stösst das Fach Biologie signifikant auf das grösste Interesse, während die Maturandinnen und Maturanden dem Fach Mathematik ebenfalls signifikant am wenigsten Interesse entgegenbringen. Diese Reihenfolge unter den drei Fächern zeigte sich auch als Ergebnis der Befragung von EVAMAR I (vgl. Ramseier et al., 2006, S. 73).

Abschliessend soll als weiterer Bereich die Studiumsvorbereitung in der Wahrnehmung der Maturandinnen und Maturanden betrachtet werden. Dazu wurden in Anlehnung an die TOSCA-Studie (vgl. Köller, Watermann, Trautwein & Lüdtke, 2004) drei Skalen zur Studienpropädeutik erhoben. Die fünf Items der ersten Skala thematisieren, in welchem Ausmass im Gymnasium Tätigkeiten ausgeführt werden, die auch für das Studium typisch sind, z. B. Informationen sammeln, Texte am Computer schreiben oder elektronische Medien zur Informationsbeschaffung nutzen (Antwortmöglichkeiten 1 = nie bis 4 = häufig). Diese Skala konnte mit einer zufrieden stellenden internen Konsistenz von Cronbachs Alpha = .72 repliziert werden. Der Mittelwert der Skala beträgt M = 2.86 (SD = .56), liegt damit über dem theoretischen Mittelwert und deutet darauf hin, dass diese Tätigkeiten im Durchschnitt am ehesten „manchmal" durchgeführt werden. Am häufigsten werden den Schülerangaben zufolge in der Schule elektronische Medien zur Informationsbeschaffung genutzt (M = 3.43, SD = .82), während eher selten Abbildungen am Computer erstellt werden (M = 2.23, SD = .87).

Die zweite Skala thematisiert weitere studienvorbereitende Tätigkeiten wie die Planung wissenschaftlicher Experimente und die Anfertigung von Protokollen oder Referaten. Diese Skala konnte allerdings nicht repliziert werden. Die Betrachtung auf der Ebene der Einzel-Items ergibt, dass am häufigsten Referate vorgetragen wurden (M = 3.64, SD = .57), während eher selten wissenschaftliche Experimente geplant (M = 1.91, SD = .89) oder Protokolle zu Diskussionen angefertigt wurden (M = 1.92, SD = .92).

Die dritte Dimension thematisiert die subjektive Beurteilung der Studienpropädeutik. Es wurde gefragt, wie die Maturandinnen und Maturanden die Vorbereitung auf verschiedene Anforderungen im Studium durch ihr Gymnasium beurteilen (z. B. „Ich glaube, ich habe im Gymnasium verstanden, wie man Fragestellungen wissenschaftlich bearbeitet" oder „Das Gymnasium hat mir viel Allgemeinwissen vermittelt"). Die Skala weist eine nicht ganz zufrieden stellende, aber noch akzeptable interne Konsistenz von .68 auf. Insgesamt ergibt sich bei einer vierstufigen Antwortskala (1 = trifft nicht zu bis 4 = trifft völlig zu) ein Mittelwert von 2.89 (SD = .43), der auf eine eher positive Beurteilung der Studienvorbereitung schliessen lässt. So weist das Item „Durch das Gymnasium fühle ich mich insgesamt gut für ein Studium vorbereitet" einen Mittelwert von 2.78 (SD = .73) aus, was ebenfalls auf eine überwiegend positive Einstellung gegenüber der studiumsvorbereitenden Funktion der Schule deutet. Am kritischsten wird die Aussage „Die Auseinandersetzung mit den Lerninhalten am Gymnasium hat mir die Studienfachwahl erleichtert" gesehen, hier beträgt der Mittelwert lediglich M = 2.34 bei einer allerdings vergleichsweise hohen Standardabweichung von SD = .98.

5.8 Interpretation der Ergebnisse im Hinblick auf die Studierfähigkeit

Die Schweizer Maturandinnen und Maturanden erzielten in den untersuchten Bereichen Ergebnisse, die für standardisierte, für eine bestimmte Population entwickelte Tests normal sind. Zu dieser Normalität gehört, dass im Mittel die Testaufgaben etwa zur Hälfte richtig gelöst wurden. Insofern kann der Ausbildungsstand der Schweizer Maturandinnen und Maturanden als zufrieden stellend bewertet werden. Die Resultate waren in Erstsprache etwas besser als in Mathematik und Biologie. In Biologie fielen die Ergebnisse für die Gruppe der Maturandinnen und Maturanden, die dieses Fach bereits ein halbes Jahr zuvor oder noch eher abgeschlossen hatten, klar schlechter aus, vermutlich als Folge des Vergessenseffekts. Dies deutet darauf hin, dass das an Maturitätsschulen erworbene Wissen nur für kurze Zeit vollumfänglich präsent ist und dass sowohl ein möglichst später Abschluss des Faches am Gymnasium wie auch eine rasche Aufnahme des Studiums vorteilhaft sind. Die Ergebnisse des überfachlichen Fähigkeitstests (UFT) liegen nur leicht unter den Resultaten des jährlich durchgeführten Medizin-Eignungstests. Das ist plausibel, weil bei den EVAMAR-Erhebungen auch alle jene Personen teilgenommen haben, die sich ein Medizinstudium wegen mangelnden Fähigkeiten nicht zutrauen. Dieses Ergebnis ist zudem ein wichtiger Indikator für die hohe Anstrengungsbereitschaft der Maturandinnen und Maturanden bei der Bearbeitung der EVAMAR-Tests. Aufgrund der Nähe der Ergebnisse des UFT zu den Resultaten, wie sie im individualbiografisch hoch relevanten Kontext der Eignungsprüfung erreicht werden, braucht die Anstrengungsbereitschaft deshalb auch den Vergleich mit jener in den gewohnten notenwirksamen Testsituationen im Gymnasium keineswegs zu scheuen.

Sowohl zwischen Einzelpersonen als auch zwischen ganzen Klassen bestehen hinsichtlich der Ergebnisse jedoch beträchtliche Unterschiede, die grösser ausfallen als bei den effektiv erteilten Maturanoten. Es lässt sich somit feststellen, dass erhebliche Anteile von Maturandinnen und Maturanden in den mit den Tests erfassten Bereichen über vermutlich ungenügende Kompetenzen in mindestens einem Testbereich verfügen. Diese Erkenntnis zeichnet sich jedoch auch schon bei einer Analyse der durch die Schulen selbst vergebenen Maturanoten ab, am ausgeprägtesten in Mathematik. 41.4% der Maturandinnen und Maturanden wurden im Jahre 2007 in der schriftlichen Maturaprüfung mit einer ungenügenden Note bewertet. Bei der Maturaendnote für Mathematik, die auch die Erfahrungsnoten und die mündliche Prüfung berücksichtigt, waren es immer noch 24.4%. Zwar bleibt offen, ob es sich in den Maturitätsprüfungen bei den für die ungenügenden Ergebnisse ursächlichen Teilen um Wissen und Können handelt, das für die allgemeine Studierfähigkeit unabdingbar ist. Im Extremfall könnten einerseits Maturandinnen und Maturanden mit ungenügenden Noten trotzdem in für die Studierfähigkeit relevanten Bereichen genügende Werte aufweisen. Andererseits könnte es auch Maturandinnen und Maturanden mit genügenden Noten geben, die in für die allgemeine Studierfähigkeit relevanten Bereichen des Fachs trotzdem über ungenügende Kompetenzen verfügen. Die teilweise schlechten Ergebnisse in den Tests hingegen betreffen durchwegs Teile, die für mindestens ein Studienfach relevant sind. Ab welcher Grenze bei den Leistungstests ein erfolgreiches Studium nicht mehr möglich ist, kann nicht deterministisch bestimmt werden. Dazu wären weitere, umfangreiche und langwierige empirische Untersuchungen notwendig. Die Vermutung aber, dass nicht alle Maturandinnen und Maturanden in der ganzen Breite über genügende Kompetenzen verfügen, um jedes beliebige Studium aufnehmen zu können, ohne zusätzlich Lücken füllen zu müssen, lässt sich – unter Einbezug der Maturanoten – auf gute Gründe stützen. Dies, obwohl sie durch die Verleihung des Matura-Ausweises als „allgemein studierfähig“ qualifiziert wurden.

Die Gymnasien entlassen also nicht alle ihre Maturandinnen und Maturanden mit Kompetenzen, die in der ganzen Breite als mindestens genügend eingeschätzt werden können. Das ist die Bestätigung eines Befundes, der aufgrund der effektiv an den Gymnasien vergebenen ungenügenden Noten an Maturitätsprüfungen vor allem in den Bereichen Mathematik und schriftliche Erstsprachprüfung schon bisher vermutet werden konnte. Das bedeutet aber auch, dass trotz gutem Ausbildungsstand Verbesserungen grundsätzlich möglich sind. Das ist keine aussergewöhnliche Schlussfolgerung, lässt sich eine solche doch für jede Schulstufe nach einer Leistungsmessung mit standardisierten Tests ziehen. Zwingender Handlungsbedarf würde dann bestehen, wenn die Aufgaben zu klar weniger als der Hälfte gelöst und der Durchschnitt der geschätzten Personenfähigkeiten deutlich unter einem mittleren Wert liegen würden. Das ist bei den Schweizer Maturandinnen und Maturanden nicht der Fall.

Interpretiert man allgemeine Studierfähigkeit dahingehend, dass jeder Maturitätsausweis genügende Eingangskompetenzen für alle möglichen Studienfächer bescheinigen sollte, so ist dies vermutlich für einen beachtlichen Teil der Maturandinnen und Maturanden nicht gegeben. EVAMAR II stellt das zwar nur für die getesteten Bereiche fest, das Bild würde sich aber vermutlich bei Einbezug weiterer Bereiche kaum ändern. Auch dieser Befund ist nicht neu, drücken sich doch diese Lücken ebenso in den Maturanoten aus, ermöglicht durch das Notenkompensationssystem des MAR. Damit sind auch die (subjektiven) Klagen von Universitäts-Dozierenden über Lücken vor allem in Erstsprache und Mathematik (siehe Kapitel II, Abschnitt 3.3) sowie die teilweise erheblichen Durchfallquoten in den ersten Prüfungen zumindest teilweise erklärbar.

Die Ergebnisse der Zusatzbefragung bestätigen bisherige Evaluationsuntersuchungen und auch die Ergebnisse der Dozierendenbefragung in Teilprojekt A4 (siehe Kapitel II) insofern, als die Fähigkeit zu selbstständigem Lernen und Arbeiten in der Selbsteinschätzung als teilweise lückenhaft beschrieben wird.

Die Grundfrage ist deshalb, ob sich künftig eine „lückenlose Studierfähigkeit“ aller Maturandinnen und Maturanden erreichen lässt. Dazu müssten sicher die Bestehensnormen an den Maturaprüfungen insofern revidiert werden, als ungenügende Noten nicht mehr zugelassen würden. Das hätte aber vermutlich zwei unerwünschte Effekte, nämlich eine Senkung der Maturitätsquote und eine teilweise Anpassung der Notengebung. Eine „lückenlose Studierfähigkeit“ für alle Maturandinnen und Maturanden bleibt deshalb wohl eine Fiktion.

6 Vergleichende Auswertungen

6.1 Grundsätzliches

Im heterogenen System des Schweizer Gymnasiums sind Vergleiche zwischen verschiedenen Gruppen von Gymnasien sowie Maturandinnen und Maturanden von besonderem Interesse. Diese werden im Folgenden angestellt.

Für die Vergleiche werden ausschliesslich die aus der „Rasch-Modellierung“ berechneten Personenfähigkeiten verwendet, die sich jeweils auf alle Testhefte beziehen (Darstellungsvariante P). Im Unterschied zur bisherigen Berechnung der Personenfähigkeiten sind diese nun jeweils zusätzlich auf einen gesamtschweizerischen Mittelwert von 500 standardisiert. Im Bereich zwischen 400 und 600 Punkten liegen etwa zwei Drittel (68.2%) aller Maturandinnen und Maturanden, zwischen 300 und 700 Punkten 95.4%, und nur 2.3% erreichen weniger als 300 oder mehr als 700 Punkte; fast niemand erzielt weniger als 200 oder mehr als 800 Punkte. Durch diese Standardisierung können Unterschiede über verschiedene Tests hinweg besser vergleichbar gemacht werden.

Für die Vergleiche von mehr als zwei Gruppen wurde jeweils ein A-posteriori-Test für multiple Mittelwertvergleiche auf Signifikanz nach Duncan durchgeführt. Bei nur zwei Gruppen wurde die Signifikanz der Unterschiede mit einem t-Test untersucht. Wegen der grossen Stichprobenumfänge können auch kleine Mittelwertsunterschiede mit wenig Praxisrelevanz signifikant sein. Deshalb erfolgte für die Vergleiche die Berechnung des Masses „d“ der Effektstärke nach Cohen (Bortz & Döring, 2006), und zwar für die jeweils grösste Mittelwertsdifferenz (d max). Es wird berechnet als Verhältnis von Mittelwertsdifferenz zur mittleren Standardabweichung der beiden verglichenen Gruppen. Die Interpretation des Effektstärkemasses lautet häufig wie folgt (Bortz & Döring, 2006):

- d = 0.20: kleiner Effekt
- d = 0.50: mittlerer Effekt
- d = 0.80: grosser Effekt

Die Gruppen, die sich jeweils signifikant unterscheiden (mindestens auf dem 5%-Niveau bzw. alpha = 0.05), sind in den Tabellen der folgenden Abschnitte aufsteigend mit Zahlen beschriftet. Die Zahl „1“ bezeichnet dabei die Gruppe mit den niedrigsten Werten. Die Zugehörigkeit zu zwei Gruppen bedeutet ein Fehlen signifikanter Unterschiede zu den anderen Mitgliedern dieser zwei Gruppen.

Tiefere Testwerte dürfen keinesfalls mit schlechterer Unterrichtsqualität an den entsprechenden Gymnasien gleichgesetzt werden, sondern haben in der Regel strukturelle Ursachen!

6.2 Vergleich zwischen den Straten (Gruppen von Gymnasien)

6.2.1 Vergleichende Testergebnisse für die Straten

6.2.1.1 Erstsprache

Die Ergebnisse des Stratenvergleichs im Erstsprachetest finden sich in den Tabellen IV.14 bis IV.18.

Tab. IV.14: Vergleich der Ergebnisse in Erstsprache zwischen den Straten: Totalergebnis (TOT_SPRACHE)

Straten	N (gewichtet)	Signifikante Untergruppen		
		1	2	3
Deutschschweiz MD3	1844	489		
Zürich	1921		498	
Romandie 1 (MD3)	3187		501	
Deutschschweiz gross	4442		501	
Deutschschweiz klein	923		502	
Romandie 2	1041			524

d max = 0.41

Tab. IV.15: Vergleich der Ergebnisse in Erstsprache zwischen den Straten: Bereich „Allgemeines Leseverstehen" (ALLG_LV)

Straten	N (gewichtet)	Signifikante Untergruppen		
		1	2	3
Deutschschweiz MD3	1844	495		
Deutschschweiz gross	4442	498		
Romandie 1 (MD3)	3187	499		
Zürich	1921	500		
Deutschschweiz klein	923		507	
Romandie 2	1041			514

d max = 0.24

Tab. IV.16: Vergleich der Ergebnisse in Erstsprache zwischen den Straten: Bereich „Detailliertes Leseverstehen" (DETAIL_LV)

Straten	N (gewichtet)	Signifikante Untergruppen				
		1	2	3	4	5
Deutschschweiz MD3	1844	487				
Romandie 1 (MD3)	3187		494			
Zürich	1921			502		
Deutschschweiz gross	4442			504	504	
Deutschschweiz klein	923				510	
Romandie 2	1041					519

d max = 0.38

Tab. IV.17: Vergleich der Ergebnisse in Erstsprache zwischen den Straten: Bereich „Sprachreflexion/Wortschatz“ (WORTS)

Straten	N (gewichtet)	Signifikante Untergruppen		
		1	2	3
Deutschschweiz MD3	1844	492		
Deutschschweiz gross	4442	493		
Zürich	1921		499	
Deutschschweiz klein	923		501	
Romandie 1 (MD3)	3187			512
Romandie 2	1041			517

d max = 0.34

Tab. IV.18: Vergleich der Ergebnisse in Erstsprache zwischen den Straten: Bereich „Sprachreflexion/Grammatik“ (GRAM)

Straten	N (gewichtet)	Signifikante Untergruppen		
		1	2	3
Deutschschweiz MD3	1844	489		
Deutschschweiz gross	4442		500	
Romandie 1 (MD3)	3187		500	
Zürich	1921		502	
Deutschschweiz klein	923			509
Romandie 2	1041			512

d max = 0.32

In Erstsprache schnitten die dreijährigen Deutschschweizer Gymnasien im Mittel signifikant am schlechtesten ab (siehe Tabelle IV.14). Die Effektstärken bei einem Vergleich mit der jeweils besten Gruppe, Romandie 2, liegen zwischen gering und mittel. Auf der Ebene der einzelnen Kompetenzbereiche ergeben sich die folgenden Differenzierungen:

- Beim „Allgemeinen Leseverstehen“ schnitten nur noch die kleinen Deutschschweizer Kantone und die Gruppe „Romandie 2“ besser ab als die Straten mit den tiefsten Werten.
- Beim „Detaillierten Leseverstehen“ ergibt sich eine nochmals differenziertere Gruppierung. Hier bildet auch das Stratum „Romandie 1 (MD3)“ eine eigene Gruppe am unteren Ende.
- Im Bereich „Sprachreflexion/Wortschatz“ fallen auch die grossen Deutschschweizer Kantone in die unterste Gruppe.
- Die Verteilung im Bereich „Grammatik“ bestätigt im Wesentlichen das Bild des Gesamtergebnisses. Allerdings gehören hier auch die kleinen Deutschschweizer Kantone zur Gruppe der Besten.

Insgesamt schnitten die Gymnasien mit nur dreijähriger Mindest-Dauer zumindest innerhalb des gleichen Sprachbereichs eher schlechter ab.

6.2.1.2 Mathematik

Die Ergebnisse für Mathematik finden sich in den Tabellen IV.19 bis IV.23.

Tab. IV.19: Vergleich der Ergebnisse in Mathematik zwischen den Straten: Totalergebnis (TOT_MATH)

Straten	N (gewichtet)	Signifikante Untergruppen				
		1	2	3	4	5
Romandie 1 (MD3)	3180	477				
Deutschschweiz MD3	1865		492			
Deutschschweiz gross	4446		496	496		
Zürich	2050			501	501	
Deutschschweiz klein	932				505	
Romandie 2	1048					522

d max = 0.48

Tab. IV.20: Vergleich der Ergebnisse in Mathematik zwischen den Straten: Bereich „Analytisches" (ANALYSIS)

Straten	N (gewichtet)	Signifikante Untergruppen				
		1	2	3	4	5
Romandie 1 (MD3)	3180	483				
Deutschschweiz MD3	1865		492			
Deutschschweiz gross	4446		496	496		
Zürich	2050			499		
Deutschschweiz klein	932				505	
Romandie 2	1048					520

d max = 0.41

Tab. IV.21: Vergleich der Ergebnisse in Mathematik zwischen den Straten: Bereich „Geometrisches" (GEOMETRIE)

Straten	N (gewichtet)	Signifikante Untergruppen			
		1	2	3	4
Romandie 1 (MD3)	3180	488			
Deutschschweiz gross	4446		494		
Deutschschweiz MD3	1865		498	498	
Zürich	2050			501	
Deutschschweiz klein	932			503	
Romandie 2	1048				514

d max = 0.28

Tab. IV.22: Vergleich der Ergebnisse in Mathematik zwischen den Straten: Bereich „Stochastisches“ (STOCHASTIK)

Straten	N (gewichtet)	Signifikante Untergruppen			
		1	2	3	4
Romandie 1 (MD3)	3180	482			
Deutschschweiz MD3	1865		492		
Zürich	2050			498	
Deutschschweiz klein	932			501	
Deutschschweiz gross	4446			501	
Romandie 2	1048				528

d max = 0.55

Tab. IV.23: Vergleich der Ergebnisse in Mathematik zwischen den Straten: Bereich „Diverses“ (VERSCH)

Straten	N (gewichtet)	Signifikante Untergruppen				
		1	2	3	4	5
Romandie 1 (MD3)	3180	477				
Deutschschweiz MD3	1865		489			
Deutschschweiz klein	932			498		
Deutschschweiz gross	4446			500	500	
Zürich	2050				505	
Romandie 2	1048					517

d max = 0.47

In Mathematik schnitten die Gymnasien mit nur dreijähriger Mindest-Dauer im Gesamtergebnis klar schlechter ab als jene mit mindestens vier Jahren, insbesondere in der Romandie. Die Effektstärken zwischen den Straten mit dem tiefsten (Romandie 1 (MD3)) und dem höchsten Wert (Romandie 2) sind mittlere. Dieses Bild wird insbesondere in den Bereichen „Stochastik“ und „Verschiedenes“ mit noch leicht höheren Effektstärken akzentuiert und umgekehrt im Bereich „Geometrie“ etwas entschärft. Es scheint sich zu zeigen, dass die Schuldauer im Gymnasium in Mathematik eine noch grössere positive Auswirkung auf den Leistungsstand hat als in Erstsprache.

6.2.1.3 Biologie

Die Vergleichsergebnisse für Biologie finden sich in den Tabellen IV.24 bis IV.30.

Tab. IV.24: Vergleich der Ergebnisse in Biologie zwischen den Straten: Totalergebnis (TOT_BIO)

Straten	N (gewichtet)	Signifikante Untergruppen			
		1	2	3	4
Romandie 1 (MD3)	3152	459			
Romandie 2	1048	462			
Deutschschweiz MD3	1884		506		
Deutschschweiz klein	916		510	510	
Deutschschweiz gross	4464			514	
Zürich	1974				522

d max = 0.75

Tab. IV.25: Vergleich der Ergebnisse in Biologie zwischen den Straten: Bereich „Evolution" (B_EVO)

Straten	N (gewichtet)	Signifikante Untergruppen			
		1	2	3	4
Romandie 1 (MD3)	3152	462			
Romandie 2	1048	464			
Deutschschweiz MD3	1884		506		
Deutschschweiz klein	916			512	
Deutschschweiz gross	4464			514	514
Zürich	1974				519

d max = 0.74

Tab. IV.26: Vergleich der Ergebnisse in Biologie zwischen den Straten: Bereich „Struktur und Funktion" (B_STRU)

Straten	N (gewichtet)	Signifikante Untergruppen			
		1	2	3	4
Romandie 1 (MD3)	3152	463			
Romandie 2	1048	463			
Deutschschweiz MD3	1884		505		
Deutschschweiz klein	916		509		
Deutschschweiz gross	4464			515	
Zürich	1974				522

d max = 0.74

Tab. IV.27: Vergleich der Ergebnisse in Biologie zwischen den Straten: Bereich „Genetik“ (B_GEN)

Straten	N (gewichtet)	Signifikante Untergruppen			
		1	2	3	4
Romandie 2	1048	463			
Romandie 1 (MD3)	3152	464			
Deutschschweiz MD3	1884		505		
Deutschschweiz klein	916		509	509	
Deutschschweiz gross	4464			513	
Zürich	1974				521

d max = 0.76

Tab. IV.28: Vergleich der Ergebnisse in Biologie zwischen den Straten: Bereich „Informationsverarbeitung“ (B_INFO)

Straten	N (gewichtet)	Signifikante Untergruppen			
		1	2	3	4
Romandie 1 (MD3)	3152	462			
Romandie 2	1048	463			
Deutschschweiz MD3	1884		504		
Deutschschweiz klein	916			509	
Deutschschweiz gross	4464			513	
Zürich	1974				522

d max = 0.75

Tab. IV.29: Vergleich der Ergebnisse in Biologie zwischen den Straten: Bereich „Ökologie“ (B_OEKO)

Straten	N (gewichtet)	Signifikante Untergruppen			
		1	2	3	4
Romandie 1 (MD3)	3152	460			
Romandie 2	1048	463			
Deutschschweiz MD3	1884		506		
Deutschschweiz klein	916		511	511	
Deutschschweiz gross	4464			513	
Zürich	1974				521

d max = 0.75

Tab. IV.30: Vergleich der Ergebnisse in Biologie zwischen den Straten: Bereich „Stoffwechsel" (B_STOFF)

Straten	N (gewichtet)	Signifikante Untergruppen			
		1	2	3	4
Romandie 1 (MD3)	3152	461			
Romandie 2	1048	462			
Deutschschweiz MD3	1884		505		
Deutschschweiz klein	916		510	510	
Deutschschweiz gross	4464			514	
Zürich	1974				522

d max = 0.75

In Biologie liegt Zürich klar an der Spitze, mit einer hohen Effektstärke im Vergleich zum Stratum mit dem tiefsten Wert. Am unteren Ende gruppieren sich die Schulen aus der Romandie. Die Unterschiede zwischen den Gymnasien mit drei- und vierjähriger Mindest-Dauer auf der Sekundarstufe II sind innerhalb der Romandie unbedeutend, in der Deutschschweiz hingegen nach wie vor grösstenteils vorhanden, und zwar zugunsten einer vierjährigen Schuldauer.

Wird der Stratenvergleich für jede der in der Gesamtergebnisdarstellung unterschiedenen beiden Hauptgruppen vorgenommen, ergibt sich folgendes aus den Tabellen IV.31 und IV.32 ersichtliche Bild: Die Reihenfolge bleibt in der Hauptgruppe mit abgeschlossenem Biologieunterricht in etwa dieselbe, die Effektstärken fallen sogar noch höher aus, wobei der Vergessenseffekt die Unterschiede vermutlich noch akzentuiert. War hingegen der Biologieunterricht noch nicht abgeschlossen, verringern sich die Unterschiede zwischen den Straten auf einem insgesamt wesentlich höheren Niveau beträchtlich. Es handelt sich hier im Wesentlichen um die Maturandinnen und Maturanden, die Biologie entweder im Schwerpunkt- oder im Ergänzungsfach belegen. Es lässt sich aber immer noch ein mittleres Gefälle zwischen „West" und „Ost" feststellen.

Tab. IV.31: Vergleich der Ergebnisse jener Personen in Biologie, die das Fach bereits abgeschlossen hatten, zwischen den Straten: Totalergebnis (TOT_BIO)

Straten	N (gewichtet)	Signifikante Untergruppen				
		1	2	3	4	5
Romandie 1 (MD3)	2293	436				
Romandie 2	840		447			
Deutschschweiz MD3	952			493		
Deutschschweiz klein	718			495		
Deutschschweiz gross	3409				502	
Zürich	1478					511

d max = 0.97

Tab. IV.32: Vergleich der Ergebnisse jener Personen in Biologie, die das Fach nicht abgeschlossen hatten, zwischen den Straten: Totalergebnis (TOT_BIO)

Straten	N (gewichtet)	Signifikante Untergruppen	
		1	2
Deutschschweiz MD3	932	519	
Romandie 2	207	520	
Romandie 1 (MD3)	858	522	
Zürich	495		554
Deutschschweiz gross	1054		554
Deutschschweiz klein	198		565

d max = 0.56

6.2.2 Vergleich der Maturanoten zwischen den Straten

Die in den Vergleichen der Fachtestergebnisse zwischen den Straten eruierten mittleren bis grossen Unterschiede werfen die Frage auf, ob sich solche auch in den effektiv vergebenen Maturanoten in den drei Fächern zeigen (siehe Tabellen IV.33 bis IV.35).

Tab. IV.33: Vergleich der Matura-Gesamtnote in Erstsprache zwischen den Straten

Straten	N (gewichtet)	Signifikante Untergruppen				
		1	2	3	4	5
Romandie 1 (MD3)	3151	4.48				
Romandie 2	1043		4.55			
Deutschschweiz MD3	1628			4.63		
Deutschschweiz gross	4514				4.67	
Zürich	2091					4.73
Deutschschweiz klein	959					4.76

d max = 0.50

Die Maturanote in Erstsprache unterscheidet sich zwischen den Straten ebenfalls signifikant, die Effektstärken zwischen den Extremwerten erweisen sich als mittelgross. Die Rangierung aber ist im Vergleich zu jener bei den Testergebnissen in Erstsprache eine andere. Eine plausible Erklärung findet sich nicht so leicht. Sicher gehört zu den möglichen Gründen, dass ein gymnasiales Erstsprache-Curriculum weit mehr umfasst und zum Teil anderes beinhaltet, als mit den Tests gemessen wurde. Es ist augenfällig, dass man in der Westschweiz über die Sprachgrenzen hinweg (Romandie und Deutschschweiz MD3) in Erstsprache tiefere Noten vergibt als in der übrigen Schweiz.

Tab. IV.34: Vergleich der Matura-Gesamtnote in Mathematik zwischen den Straten

Straten	N (gewichtet)	Signifikante Untergruppen	
		1	2
Romandie 1 (MD3)	3151	4.29	
Deutschschweiz MD3	1628	4.30	
Deutschschweiz gross	4514	4.35	4.35
Romandie 2	1043		4.38
Deutschschweiz klein	959		4.39
Zürich	2091		4.39

d max = 0.12

Bei der Maturanote in Mathematik ist die Differenz zwischen dem tiefsten und höchsten Mittelwert zwar signifikant, beträgt aber nur eine Zehntelnote. Die Effektstärke ist hier sehr gering.

Tab. IV.35: Vergleich der Maturanote in Biologie (Erfahrungsnote) zwischen den Straten

Straten	N (gewichtet)	Signifikante Untergruppen			
		1	2	3	4
Zürich	2091	4.58			
Deutschschweiz klein	826		4.63		
Deutschschweiz MD3	1471		4.64		
Romandie 2	1043			4.70	
Deutschschweiz gross	4507			4.73	
Romandie 1 (MD3)	1172				4.77

d max = 0.36

Tab. IV.36: Vergleich der Matura-Gesamtnote in Naturwissenschaften zwischen den Straten

Straten	N (gewichtet)	Signifikante Untergruppen		
		1	2	3
Romandie 1 (MD3)	3151	4.52		
Zürich	2067	4.54		
Deutschschweiz MD3	1628		4.60	
Deutschschweiz klein	964		4.61	
Romandie 2	1043		4.63	4.63
Deutschschweiz gross	4514			4.66

d max = 0.26

Interessant ist die im Vergleich zu den Testergebnissen im Grossen und Ganzen umgekehrte Reihenfolge der besten und schlechtesten Straten bei der Erfahrungsnote Biologie mit immerhin einer beinahe mittleren Effektstärke. Auch bei der Maturanote in Naturwissenschaften (vgl. Tabelle IV.36) ergeben sich zwischen den Straten nur geringfügige Unterschiede.

Insgesamt kann festgestellt werden, dass bei den Maturanoten keine oder nur noch kleine Unterschiede zwischen den Straten zu finden sind. Vermutlich lässt sich dieser Befund damit erklären, dass Bewertungen an der Matura nicht an einer gesamtschweizerischen Norm ausgerichtet werden (können).

6.2.3 Vergleich der überfachlichen Fähigkeiten zwischen den Straten

In den nächsten fünf Tabellen (IV.37 bis IV.40) findet sich ein Vergleich der getesteten überfachlichen Fähigkeiten zwischen den Deutschschweizer Straten. Die französischsprachigen Maturandinnen und Maturanden wurden in diesem Bereich nicht getestet (Begründung siehe Abschnitt 1.1).

Tab. IV.37: Vergleich der Ergebnisse des überfachlichen Fähigkeitstests zwischen den Straten: Bereich „Quantitative und formale Probleme“

Straten	N (gewichtet)	Signifikante Untergruppen		
		1	2	3
Deutschschweiz MD3	1821	487		
Deutschschweiz gross	4471		496	
Deutschschweiz klein	933			502
Zürich	2053			506

d max = 0.23

Tab. IV.38: Vergleich der Ergebnisse des überfachlichen Fähigkeitstests zwischen den Straten: Bereich „Diagramme und Tabellen“

Straten	N (gewichtet)	Signifikante Untergruppen			
		1	2	3	4
Deutschschweiz MD3	1821	487			
Deutschschweiz gross	4471		495		
Deutschschweiz klein	933			501	
Zürich	2053				507

d max = 0.25

Tab. IV.39: Vergleich der Ergebnisse des überfachlichen Fähigkeitstests zwischen den Straten: Bereich „Textverständnis“

Straten	N (gewichtet)	Signifikante Untergruppen		
		1	2	3
Deutschschweiz MD3	1821	486		
Deutschschweiz gross	4471		496	
Deutschschweiz klein	933		501	
Zürich	2053			507

d max = 0.24

Tab. IV.40: Vergleich der Ergebnisse des überfachlichen Fähigkeitstests zwischen den Straten: Bereich „Naturwissenschaftliches Grundverständnis“

Straten	N (gewichtet)	Signifikante Untergruppen		
		1	2	3
Deutschschweiz MD3	1821	486		
Deutschschweiz gross	4471		496	
Deutschschweiz klein	933		500	
Zürich	2053			507

d max = 0.25

Auch im überfachlichen Fähigkeitstest schneidet die Gruppe der Gymnasien mit dreijähriger Mindest-Dauer wie bei den meisten Fachtestbereichen in allen Unterbereichen signifikant schlechter ab. Die Effektstärken zwischen den Extremgruppen liegen allerdings auf einem tiefen Niveau. Vermutlich bestätigt sich hier der Anspruch des Tests, Fähigkeiten zu prüfen, die relativ unabhängig vom Umfang des in den verschiedenen Fächern erarbeiteten Wissens sind. Die Schuldauer am Gymnasium scheint aber trotzdem bedeutsam zu bleiben, wenn auch in geringerem Ausmass.

6.2.4 Vergleich ausgewählter Ergebnisse aus der Zusatzbefragung

In den folgenden Tabellen IV.41 bis IV.43 findet sich ein Vergleich der erfragten hochschulrelevanten überfachlichen Kompetenzen. Es werden nur diejenigen aufgeführt, bei denen sich zwischen den Extremwerten Unterschiede im Ausmass von mindestens kleinen Effektstärken zeigen.

Tab. IV.41: Vergleich der Befragungsergebnisse „überfachliche Kompetenzen“ zwischen den Straten: Bereich „Zeitplanung“ (1 = schlechte Zeitplanung, 5 = sehr gute Zeitplanung)

Straten	N (gewichtet	Signifikante Untergruppen		
		1	2	3
Romandie 1 (MD3)	3153	2.55		
Zürich	2022	2.56		
Deutschschweiz MD3	1879		2.67	
Deutschschweiz klein	916		2.68	
Deutschschweiz gross	4388		2.68	
Romandie 2	1035			2.78

d max = 0.24

Tab. IV.42: Vergleich der Befragungsergebnisse „überfachliche Kompetenzen“ zwischen den Straten: Bereich „Prüfungsvorbereitung“ (1 = ineffiziente Prüfungsvorbereitung, 5 = sehr zeiteffiziente Prüfungsvorbereitung)

Straten	N (gewichtet)	Signifikante Untergruppen			
		1	2	3	4
Zürich	2022	3.56			
Deutschschweiz MD3	1879	3.58			
Deutschschweiz gross	4388		3.64		
Deutschschweiz klein	916		3.68		
Romandie 1 (MD3)	3153			3.75	
Romandie 2	1035				3.84

d max = 0.41

Tab. IV.43: Vergleich der Befragungsergebnisse „überfachliche Kompetenzen“ zwischen den Straten: Bereich „Zeiteffizienz“ (1 = sehr schlechte Zeiteffizienz, 5 = sehr gute Zeiteffizienz)

Straten	N (gewichtet)	Signifikante Untergruppen				
		1	2	3	4	5
Romandie 1 (MD3)	3153	3.62				
Romandie 2	1035		3.70			
Deutschschweiz MD3	1879			3.77		
Zürich	2022			3.79	3.79	
Deutschschweiz gross	4388				3.83	
Deutschschweiz klein	916					3.90

d max = 0.43

Zwischen den Straten zeigen sich höchstens mittlere und kaum sinnvoll interpretierbare Unterschiede, da sie sich im Bereich der gleichen Antwortstufe (zwischen 3.5 und 4) bewegen und die Reihenfolgen keine Konsistenz aufweisen.

6.3 Vergleich nach Schwerpunktfächern

Die Einführung einer im Vergleich zu den bisherigen Maturitätstypen erweiterten Zahl von Schwerpunktfächern war einer der wesentlichen Punkte der Maturitätsreform 1995 (siehe Kapitel I, Abschnitt 1). Während die Ausbildung in den obligatorischen Grundlagenfächern die allgemeine Studierfähigkeit sichern soll, ermöglicht die Wahl des Schwerpunktfachs (und später eines Ergänzungsfachs) den Gymnasiastinnen und Gymnasiasten die interessengeleitete, individualisierende Vertiefung ihrer gymnasialen Bildung. Zwar zeigen die Studie EVAMAR I (Ramseier et al., 2004) und weitere Evaluationsuntersuchungen (Oelkers, 2008), dass bei gewissen Schwerpunktfächern eine Kanalisierung der Studienwahl auf bestimmte Studienfächer erfolgt, aber dennoch sollten Maturandinnen und Maturanden im Sinne einer allgemeinen Studierfähigkeit in der Lage sein, unabhängig von der Wahl des Schwerpunkt- und Ergänzungsfachs jedes Hochschulfach studieren zu können. Gemäss den gleichen Studien, die auf Befragungen basieren, unterscheiden sich Maturandinnen und Maturanden mit verschiedenen Schwerpunktfächern nicht hinsichtlich ihres Wissens und Könnens in den Grundlagenfächern (Ramseier et al., 2004, 138; Oelkers, 2008, 120). Vor diesem Hintergrund ist es von grossem Interesse, ob sich dieser Befund auch in den Leistungsdaten spiegelt, die im Rahmen des Projekts EVAMAR II erhoben wurden.

6.3.1 Vergleich der Fach-Testergebnisse nach Schwerpunktfächern

6.3.1.1 Erstsprache

Im Folgenden werden die Testergebnisse in Erstsprache der Maturandinnen und Maturanden mit verschiedenen Schwerpunktfächern miteinander verglichen (Tabellen IV.44 bis IV.48).

Tab. IV.44: Vergleich der Ergebnisse des Tests in Erstsprache zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Gesamtergebnis (TOT_SPRACHE)

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen				
		1	2	3	4	5
Bildnerisches Gestalten	1176	477				
Wirtschaft und Recht	3097		492			
Musik	762		494			
Physik und Anwendungen der Mathematik	1403			501		
Moderne Sprachen	3217			502		
Biologie und Chemie	1722				510	
Philosophie/Pädagogik/ Psychologie	1312				514	
Alte Sprachen	669					533

d max = 0.59

Tab. IV.45: Vergleich der Ergebnisse des Tests in Erstsprache zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Allgemeines Leseverstehen" (ALLG_LV)

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen			
		1	2	3	4
Bildnerisches Gestalten	1176	484			
Wirtschaft und Recht	3097	489			
Musik	762		495		
Moderne Sprachen	3217		501		
Physik und Anwendungen der Mathematik	1403		501		
Philosophie/Pädagogik/ Psychologie	1312			508	
Biologie und Chemie	1722			510	
Alte Sprachen	669				536

d max = 0.63

Tab. IV.46: Vergleich der Ergebnisse des Tests in Erstsprache zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Detailliertes Leseverstehen" (DETAIL_LV)

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen					
		1	2	3	4	5	6
Bildnerisches Gestalten	1176	485					
Wirtschaft und Recht	3097		492				
Musik	762		493				
Physik und Anwendungen der Mathematik	1403		498	498			
Biologie und Chemie	1722			503	503		
Moderne Sprachen	3217				507	507	
Philosophie/Pädagogik/ Psychologie	1312					511	
Alte Sprachen	669						526

d max = 0.46

Tab. IV.47: Vergleich der Ergebnisse des Tests in Erstsprache zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Sprachreflexion/Wortschatz“ (WORTS)

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen				
		1	2	3	4	5
Musik	762	484				
Bildnerisches Gestalten	1176	485				
Wirtschaft und Recht	3097		491			
Moderne Sprachen	3217			503		
Physik und Anwendungen der Mathematik	1403			506	506	
Philosophie/Pädagogik/ Psychologie	1312				510	
Biologie und Chemie	1722				511	
Alte Sprachen	669					527

d max = 0.58

Tab. IV.48: Vergleich der Ergebnisse des Tests in Erstsprache zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Sprachreflexion/Grammatik“ (GRAM)

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen				
		1	2	3	4	5
Bildnerisches Gestalten	1176	487				
Musik	762	488				
Wirtschaft und Recht	3097	491	491			
Biologie und Chemie	1722		494			
Physik und Anwendungen der Mathematik	1403			501		
Moderne Sprachen	3217				509	
Philosophie/Pädagogik/ Psychologie	1312				511	
Alte Sprachen	669					531

d max = 0.59

Maturandinnen und Maturanden mit dem Schwerpunktfach „Alte Sprachen“ (vorwiegend Latein) haben im Erstsprachtest in allen Bereichen durchschnittlich am besten, jene mit dem Schwerpunktfach „Bildnerisches Gestalten“ ebenfalls in allen Bereichen und teilweise mit den Gruppen „Musik“ (in den Bereichen Sprachreflexion/Wortschatz und Grammatik) sowie „Wirtschaft und Recht“ (bei Allgemeinem Leseverstehen und Grammatik) am schlechtesten abgeschlossen. Die Effektstärken zwischen den Extremgruppen sind mittel. Es fällt auf, dass sich die Maturandinnen und Maturanden des Sprachen-Schwerpunkts „Moderne Sprachen“ in einem Sprachtest „nur“ in der Mitte platzierten.

6.3.1.2 Mathematik

Der Vergleich der Testergebnisse in Mathematik findet sich in den Tabellen IV.49 bis IV.53. Bei diesem Vergleich ist zu berücksichtigen, dass Maturandinnen und Maturanden des Schwerpunktfachs „Physik und Anwendungen der Mathematik“ eine vertiefte mathematische Bildung erhalten. Deshalb ist bei der Effektstärke in Klammer zusätzlich der Wert für die grösste Differenz zwischen den anderen Schwerpunktgruppen angegeben.

Tab. IV.49: Vergleich der Ergebnisse des Tests in Mathematik zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Gesamtergebnis (TOT_MATH)

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen						
		1	2	3	4	5	6	7
Bildnerisches Gestalten	1156	455						
Philosophie/Pädagogik/Psychologie	1313		462					
Musik	698		467					
Moderne Sprachen	3336			475				
Wirtschaft und Recht	3181				485			
Biologie und Chemie	1741					504		
Alte Sprachen	671						512	
Physik und Anwendungen der Mathematik	1416							614

d max = 1.96 (0.69)

Tab. IV.50: Vergleich der Ergebnisse des Tests in Mathematik zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Analytisches“ (ANALYSIS)

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen					
		1	2	3	4	5	6
Philosophie/Pädagogik/ Psychologie	1313	462					
Bildnerisches Gestalten	1156	463					
Musik	698	465					
Moderne Sprachen	3336		478				
Wirtschaft und Recht	3181			490			
Biologie und Chemie	1741				500		
Alte Sprachen	671					509	
Physik und Anwendungen der Mathematik	1416						608

d max = 1.85 (0.57)

Tab. IV.51: Vergleich der Ergebnisse des Tests in Mathematik zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Geometrisches“ (GEOMETRIE)

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen				
		1	2	3	4	5
Bildnerisches Gestalten	1156	462				
Philosophie/Pädagogik/ Psychologie	1313	464				
Musik	698		478			
Moderne Sprachen	3336		478			
Wirtschaft und Recht	3181			488		
Biologie und Chemie	1741				508	
Alte Sprachen	671				508	
Physik und Anwendungen der Mathematik	1416					608

d max = 0.92 (0.59)

Tab. IV.52: Vergleich der Ergebnisse des Tests in Mathematik zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Stochastisches“ (STOCHASTIK)

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen						
		1	2	3	4	5	6	7
Bildnerisches Gestalten	1156	463						
Musik	698		472					
Philosophie/Pädagogik/Psychologie	1313		476					
Moderne Sprachen	3336			484				
Wirtschaft und Recht	3181				491			
Biologie und Chemie	1741					501		
Alte Sprachen	671						521	
Physik und Anwendungen der Mathematik	1416							582

d max = 1.43 (0.71)

Tab. IV.53: Vergleich der Ergebnisse des Tests in Mathematik zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Diverses“ (VERSCH)

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen					
		1	2	3	4	5	6
Bildnerisches Gestalten	1156	468					
Philosophie/Pädagogik/Psychologie	1313	468					
Musik	698		479				
Moderne Sprachen	3336		481				
Wirtschaft und Recht	3181			490			
Biologie und Chemie	1741				498		
Alte Sprachen	671					510	
Physik und Anwendungen der Mathematik	1416						586

d max = 1.54 (0.55)

Maturandinnen und Maturanden des Schwerpunkts „Physik und Anwendungen der Mathematik“ haben in allen Bereichen des Mathematiktests klar am besten, die Gruppe „Bildnerisches Gestalten“ ebenfalls in allen Bereichen und teilweise zusammen mit der Gruppe „Philosophie/Pädagogik/Psychologie“ (bei den Bereichen „Analytisches“, „Geometrisches“ und „Diverses“) und dem Schwerpunkt „Musik“ (im Bereich „Analytisches“) am schlechtesten abgeschlossen. Das Resultat der besten Gruppe fällt erwartungsgemäss aus. Die Effektstärke zwischen dieser und der schlechtesten Gruppe ist aber ausserordentlich gross. Die Maturandinnen und Maturanden des Schwerpunkts „Alte Sprache“ haben am zweitbesten abgeschnitten, noch vor jenen des Schwerpunkts „Biologie und Chemie“ und jenen des Schwerpunkts „Wirtschaft und Recht“. Die Effektstärken der zweitbesten Gruppe bzw. der besten „Nicht-Spezialisten-Gruppe“ „Alte Sprachen“ zur jeweils schlechtesten liegen immer noch bei erheblichen Werten, nämlich zwischen d = 0.71 (Stochastik) und d = 0.55 (Verschiedenes).

6.3.1.3 Biologie

In den Tabellen IV.54 bis IV.60 werden die Testergebnisse in Biologie gegenübergestellt. Auch bei diesem Vergleich ist eine fachspezifische Kombination von Grundlagen- und Schwerpunktfach zu berücksichtigen. Maturandinnen und Maturanden des Schwerpunktfachs „Biologie und Chemie“ besuchen mehr Unterrichtsstunden in Biologie und sollten in diesem Fach ganz klar über das bessere Wissen und Können verfügen. Deshalb ist bei der Effektstärke wieder in Klammer zusätzlich der Wert für die grösste Differenz zwischen den anderen Schwerpunktgruppen angegeben.

Tab. IV.54: Vergleich der Ergebnisse des Tests in Biologie zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Totalergebnis (TOT_BIO)

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen				
		1	2	3	4	5
Musik	763	475				
Moderne Sprachen	3260		485			
Philosophie/Pädagogik/ Psychologie	1309		484			
Wirtschaft und Recht	3137		487			
Bildnerisches Gestalten	1167		491	491		
Physik und Anwendungen der Mathematik	1402			496		
Alte Sprachen	668				505	
Biologie und Chemie	1723					559

d max = 1.08 (0.35)

Tab. IV.55: Vergleich der Ergebnisse des Tests in Biologie zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Evolution" (B_EVO)

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen			
		1	2	3	4
Musik	763	483			
Wirtschaft und Recht	3137	486			
Moderne Sprachen	3260	486			
Philosophie/Pädagogik/ Psychologie	1309	489			
Bildnerisches Gestalten	1167		496		
Physik und Anwendungen der Mathematik	1402		498		
Alte Sprachen	668			504	
Biologie und Chemie	1723				554

d max = 0.988 (0.27)

Tab. IV.56: Vergleich der Ergebnisse des Tests in Biologie zwischen den Gruppen unterschiedlicher Schwerpunktfächer Bereich „Struktur und Funktion“ (B_STRU)

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen			
		1	2	3	4
Wirtschaft und Recht	3137	484			
Musik	763	485			
Moderne Sprachen	3260	485			
Philosophie/Pädagogik/ Psychologie	1309	487			
Bildnerisches Gestalten	1167		497		
Physik und Anwendungen der Mathematik	1402		498		
Alte Sprachen	668			506	
Biologie und Chemie	1723				557

d max = 0.99 (0.27)

Tab. IV.57: Vergleich der Ergebnisse des Tests in Biologie zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Genetik“ (B_GEN)

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen			
		1	2	3	4
Wirtschaft und Recht	3137	484			
Musik	763	484			
Moderne Sprachen	3260	486			
Philosophie/Pädagogik/ Psychologie	1309	487	487		
Bildnerisches Gestalten	1167		492		
Physik und Anwendungen der Mathematik	1402			499	
Alte Sprachen	668			502	
Biologie und Chemie	1723				557

d max = 0.99 (0.22)

Tab. IV.58: Vergleich der Ergebnisse des Tests in Biologie zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Informationsverarbeitung“ (B_INFO)

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen					
		1	2	3	4	5	6
Musik	763	481					
Moderne Sprachen	3260	484	484				
Wirtschaft und Recht	3137	484	484				
Philosophie/Pädagogik/ Psychologie	1309		488	488			
Bildnerisches Gestalten	1167			492			
Physik und Anwendungen der Mathematik	1402				499		
Alte Sprachen	668					506	
Biologie und Chemie	1723						557

d max = 1.00 (0.30)

Tab. IV.59: Vergleich der Ergebnisse des Tests in Biologie zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Ökologie“ (B_OEKO)

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen				
		1	2	3	4	5
Musik	763	484				
Wirtschaft und Recht	3137	484				
Moderne Sprachen	3260	484				
Philosophie/Pädagogik/ Psychologie	1309	488	488			
Bildnerisches Gestalten	1167		493	493		
Physik und Anwendungen der Mathematik	1402			498		
Alte Sprachen	668				506	
Biologie und Chemie	1723					555

d max = 0.95 (0.28)

Tab. IV.60: Vergleich der Ergebnisse des Tests in Biologie zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Stoffwechsel“ (B_STOFF)

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen				
		1	2	3	4	5
Wirtschaft und Recht	3137	483				
Musik	763	484				
Moderne Sprachen	3260	485				
Philosophie/Pädagogik/ Psychologie	1309	488	488			
Bildnerisches Gestalten	1167		492	492		
Physik und Anwendungen der Mathematik	1402			498		
Alte Sprachen	668				505	
Biologie und Chemie	1723					559

d max = 1.02 (0.28)

Im Biologietest haben die Maturandinnen und Maturanden des Schwerpunkts „Biologie und Chemie“ wiederum erwartungskonform in allen Bereichen die besten Ergebnisse erzielt. Auch in diesem Test liegen die Maturandinnen und Maturanden des Schwerpunkts „Alte Sprachen“ in allen Bereichen an zweiter Stelle, noch vor jenen des anderen naturwissenschaftlichen Schwerpunkts „Physik und Anwendungen der Mathematik“. Die Ergebnisse der Maturandinnen und Maturanden des Schwerpunkts „Musik“ sind im Gesamttest klar am schlechtesten ausgefallen. In den Teilbereichen befinden sie sich zusammen mit den Maturandinnen und Maturanden der Schwerpunkte „Moderne Sprachen“ (bei „Evolution“, „Struktur und Funktion“, „Genetik“, „Informationsverarbeitung“, „Ökologie“, „Stoffwechsel“), „Philosophie/Pädagogik/Psychologie“ (bei „Evolution“, „Struktur und Funktion“, „Genetik“, „Ökologie“, „Stoffwechsel“) sowie „Wirtschaft und Recht“ (bei „Evolution“, „Struktur und Funktion“, „Genetik“, „Informationsverarbeitung“, „Ökologie“, „Stoffwechsel“) in der schlechtesten Ergebnisgruppe. Die Effektstärken der jeweils schlechtesten Schwerpunkt-Gruppe zur „Spezialistengruppe“ sind sehr hoch und bleiben im Vergleich zur in allen Bereichen zweitbesten Gruppe „Alte Sprachen“ immer noch moderat, zwischen d = 0.35 (Gesamttest) und d = 0.22 (Genetik).

6.3.2 Vergleich der Maturanoten nach Schwerpunktfächern

Nachdem ein Vergleich der Testergebnisse der Schwerpunktfach-Gruppen teilweise grosse Unterschiede ergeben hat, stellt sich die Frage, ob auch die effektiv vergebenen Maturanoten in den drei Fächern so stark differieren (siehe Tabellen IV.61 bis IV.64).

Tab. IV.61: Vergleich der vergebenen Maturanote in Erstsprache zwischen den Gruppen verschiedener Schwerpunktfächer

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen				
		1	2	3	4	5
Wirtschaft und Recht	3138	4.54				
Biologie und Chemie	1713	4.58				
Bildnerisches Gestalten	1152	4.58				
Physik und Anwendungen der Mathematik	1421		4.63			
Moderne Sprachen	3344			4.67		
Philosophie/Pädagogik/ Psychologie	1207			4.67		
Musik	729				4.75	
Alte Sprachen	682					4.87

d max = 0.60

Die Effektstärke zwischen den Extremgruppen bleibt beim Vergleich der Maturanoten in Erstsprache eine mittlere (siehe Tabelle IV.61). Hervorzuheben sind die vergleichsweise zur Testrangierung verbesserten Bewertungen für die Maturandinnen und Maturanden des Schwerpunkts „Musik“ und das relative Abgleiten jener des Schwerpunkts „Biologie und Chemie“.

Tab. IV.62: Vergleich der vergebenen Maturanote in Mathematik zwischen den Gruppen verschiedener Schwerpunktfächer

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen				
		1	2	3	4	5
Philosophie/Pädagogik/ Psychologie	1207	4.14				
Bildnerisches Gestalten	1152	4.19				
Moderne Sprachen	3344		4.26			
Wirtschaft und Recht	3138		4.28			
Musik	729		4.30	4.31		
Biologie und Chemie	1713			4.34		
Alte Sprachen	682				4.57	
Physik und Anwendungen der Mathematik	1421					4.88

d max = 0.93

Beim Vergleich der Maturanoten in Mathematik halbiert sich der mit der Effektstärke gemessene Unterschied zwischen den beiden Extremgruppen gegenüber dem Testunterschied beinahe; er bleibt aber sehr hoch (siehe Tabelle IV.62). In diesem Fach stimmt auch die sonstige Rangierung recht gut mit jener der Testergebnisse überein. Offenbar gelingt es den bewertenden Personen in Mathematik eher, sich bei der Benotung implizit an objektiven Güte-Standards zu orientieren. Wieder fallen Maturandinnen und Maturanden des Schwerpunkts „Alte Sprachen“ dadurch auf, dass sie sich noch vor jenen mit dem Schwerpunkt „Biologie und Chemie“ platzieren.

Tab. IV.63: Vergleich der vergebenen Matura-Erfahrungsnote in Biologie zwischen den Gruppen verschiedener Schwerpunktfächer

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen			
		1	2	3	4
Musik	692	4.61			
Moderne Sprachen	3007	4.61			
Wirtschaft und Recht	2635	4.65	4.65		
Philosophie/Pädagogik/ Psychologie	985		4.66		
Physik und Anwendungen der Mathematik	1033		4.69		
Biologie und Chemie	1082			4.80	
Bildnerisches Gestalten	1056			4.81	
Alte Sprachen	618				4.91

d max = 0.54

Beim Vergleich der Matura-Erfahrungsnoten in Biologie fällt auf, dass die Effektstärke beim Unterschied zwischen den Extremgruppen auf ein mittleres Mass sinkt (siehe Tabelle IV.63). Die Platzierung stimmt mit kleinen Abweichungen erstaunlich gut mit der Rangierung des Biologietests überein. Die grösste Abweichung kann für die Maturandinnen und Maturanden des Schwerpunkts „Biologie und Chemie“ sowie „Bildnerisches Gestalten“ festgestellt werden. Erstere wurden mittels Noten vergleichsweise schlechter bewertet (und diesbezüglich erstaunlicherweise gar von den „Altsprachlern“ überholt), letztere erhielten bessere Bewertungen.

Tab. IV.64: Vergleich der vergebenen Maturanote in Naturwissenschaften zwischen den Gruppen verschiedener Schwerpunktfächer

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen				
		1	2	3	4	5
Musik	729	4.48				
Philosophie/Pädagogik/ Psychologie	1207	4.49				
Moderne Sprachen	3344	4.52				
Wirtschaft und Recht	3138		4.58			
Bildnerisches Gestalten	1152			4.65		
Biologie und Chemie	1691			4.65		
Physik und Anwendungen der Mathematik	1423				4.72	
Alte Sprachen	682					4.83

d max = 0.64

Obwohl sich die Matura-Gesamtnoten in Naturwissenschaften nur bedingt mit den Testergebnissen in Biologie vergleichen lassen, ist auch die Reihenfolge in dieser Darstellung interessant (siehe Tabelle IV.64). Die Effektstärke zwischen den beiden Extremgruppen ist etwas höher als beim Vergleich der Biologie-Erfahrungsnoten, aber klar tiefer als beim Vergleich der Biologie-Testergebnisse. Als deutlich beste Gruppe erweisen sich – wie schon bei der Erfahrungsnote in Biologie – die „Altsprachler“, die erstaunlicherweise selbst in den Naturwissenschaften eine bessere Bewertung erreichten als die Naturwissenschaft-Spezialisten.

Ergänzend zu diesen Mittelwertvergleichen sind aus Tabelle IV.65 die Prozentzahlen für ungenügende Ergebnisse ersichtlich. In Mathematik zeigen sich ungenügende Bewertungen in der schriftlichen Maturitätsprüfung für

die Hälfte oder mehr aller Maturandinnen und Maturanden in den nach MAR 95 neuen Schwerpunkten „Philosophie/Pädagogik/Psychologie“, „Bildnerisches Gestalten“ und „Musik“. Im Weiteren haben offenbar Erfahrungsnoten kompensierende Wirkung für die in schriftlichen Prüfungen eher hohe Zahl an ungenügenden Bewertungen.

Tab. IV.65: Vergleich des Prozentanteils ungenügender Ergebnisse in den Maturitätsprüfungen zwischen den Gruppen unterschiedlicher Schwerpunktfächer

Schwerpunkt	Schriftliche Erstsprachprüfung	Gesamt-Maturanote Erstsprache	Schriftliche Mathematikprüfung	Gesamt-Maturanote Mathematik	Biologie-Erfahrungsnote	Gesamt-Maturanote Naturwissenschaften
Alte Sprachen	9.7%	1.3%	32.3%	19.9%	6.0%	2.7%
Moderne Sprachen	19.2%	4.3%	46.7%	29.6%	8.7%	8.4%
Physik und Anwendungen der Mathematik	14.9%	4.8%	12.4%	6.3%	6.1%	2.3%
Biologie und Chemie	22.7%	6.2%	41.8%	17.2%	3.2%	3.5%
Wirtschaft und Recht	22.0%	5.8%	40.7%	25.5%	4.8%	5.2%
Philosophie/ Pädagogik/ Psychologie	20.6%	3.7%	55.8%	32.3%	3.4%	8.9%
Bildnerisches Gestalten	24.2%	5.3%	48.8%	32.3%	2.5%	3.4%
Musik	13.0%	1.3%	48.6%	25.6%	5.4%	5.9%

6.3.3 Vergleich der überfachlichen Fähigkeiten nach Schwerpunktfächern

Die nächsten vier Tabellen (IV.66 bis IV.69) zeigen einen Vergleich der getesteten überfachlichen Fähigkeiten zwischen den Gruppen unterschiedlicher Schwerpunktfächer.

Die Reihenfolge der Testergebnisse ist in allen Unterbereichen dieselbe. Die Unterschiede zwischen der besten Gruppe („Physik und Anwendungen der Mathematik“) sowie der schlechtesten („Philosophie/Pädagogik/Psychologie“) sind erheblich (sehr grosse Effektstärken). Auch die Maturandinnen und Maturanden des Schwerpunkts „Bildnerisches Gestalten“ befinden sich in der Signifikanzgruppe mit den schlechtesten Ergebnissen.

Tab. IV.66: Vergleich der Ergebnisse des überfachlichen Fähigkeitstests zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Quantitative und formale Probleme

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen				
		1	2	3	4	5
Philosophie/Pädagogik/Psychologie	834	469				
Bildnerisches Gestalten	901	473				
Musik	665		485			
Moderne Sprachen	2426		486			
Wirtschaft und Recht	2359			499		
Alte Sprachen	391				514	
Biologie und Chemie	939				517	
Physik und Anwendungen der Mathematik	761					560

d max = 1.25

Tab. IV.67: Vergleich der Ergebnisse des überfachlichen Fähigkeitstests zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Diagramme und Tabellen

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen					
		1	2	3	4	5	6
Philosophie/Pädagogik/ Psychologie	834	468					
Bildnerisches Gestalten	901	475	475				
Musik	665		481	481			
Moderne Sprachen	2426			487			
Wirtschaft und Recht	2359				499		
Alte Sprachen	391				506		
Biologie und Chemie	939					519	
Physik und Anwendungen der Mathematik	761						557

d max = 1.22

Tab. IV.68: Vergleich der Ergebnisse des überfachlichen Fähigkeitstests zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Textverständnis

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen				
		1	2	3	4	5
Philosophie/Pädagogik/ Psychologie	834	476				
Bildnerisches Gestalten	901	478				
Musik	665	483	483			
Moderne Sprachen	2426		488			
Wirtschaft und Recht	2359			496		
Alte Sprachen	391				514	
Biologie und Chemie	939				519	
Physik und Anwendungen der Mathematik	761					553

d max = 1.02

Tab. IV.69: Vergleich der Ergebnisse des überfachlichen Fähigkeitstests zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Naturwissenschaftliches Grundverständnis

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen					
		1	2	3	4	5	6
Philosophie/Pädagogik/ Psychologie	834	474					
Bildnerisches Gestalten	901	477					
Musik	665		485				
Moderne Sprachen	2426		488				
Wirtschaft und Recht	2359			496			
Alte Sprachen	391				511		
Biologie und Chemie	939					521	
Physik und Anwendungen der Mathematik	761						549

d max = 1.01

6.3.4 Vergleich ausgewählter Ergebnisse aus der Zusatzbefragung nach Schwerpunktfächern

In den folgenden Tabellen IV.70 bis IV.74 findet sich ein Vergleich der erfragten hochschulrelevanten überfachlichen Kompetenzen, wobei wieder nur diejenigen aufgeführt sind, bei denen sich zwischen den Extremwerten Unterschiede im Ausmass von mindestens kleinen Effektstärken zeigen.

Tab. IV.70: Vergleich der Befragungsergebnisse „überfachliche Kompetenzen" zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Zeitplanung" (1 = schlechte Zeitplanung, 5 = sehr gute Zeitplanung)

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen				
		1	2	3	4	5
Physik und Anwendungen der Mathematik	1418	2.32				
Biologie und Chemie	1733		2.58			
Alte Sprachen	656		2.63	2.63		
Moderne Sprachen	3304			2.66	2.66	
Wirtschaft und Recht	3145			2.69	2.69	
Musik	706			2.70	2.70	2.70
Bildnerisches Gestalten	1135				2.73	2.73
Philosophie/Pädagogik/ Psychologie	1296					2.78

d max = 0.50

Maturandinnen und Maturanden mit dem Schwerpunkt „Physik und Anwendungen der Mathematik" schätzen ihre Zeitplanung als am schlechtesten ein, jene des Schwerpunkts „Philosophie/Pädagogik/Psychologie" ihre als am besten (siehe Tabelle IV.70). Der Unterschied ist ein mittelgrosser und macht immerhin einen halben Punkt auf der vierstufigen Skala aus. Die „Physiker/Mathematiker" beurteilen auch ihre Prüfungsvorbereitungseffizienz vergleichsweise am schlechtesten (siehe Tabelle IV.71).

Tab. IV.71: Vergleich der Befragungsergebnisse „überfachliche Kompetenzen" zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Prüfungsvorbereitung" (1 = ineffiziente Prüfungsvorbereitung, 5 = sehr zeiteffiziente Prüfungsvorbereitung)

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen				
		1	2	3	4	5
Physik und Anwendungen der Mathematik	1418	3.55				
Musik	706	3.59	3.59			
Wirtschaft und Recht	3145		3.61	3.61		
Bildnerisches Gestalten	1135		3.62	3.62		
Biologie und Chemie	1733			3.64		
Philosophie/Pädagogik/ Psychologie	1296				3.72	
Moderne Sprachen	3304				3.74	
Alte Sprachen	656					3.86

d max = 0.46

Alle in den weiteren Tabellen dargestellten Unterschiede sind zwar signifikant, aber gering und auf der Skala minimal.

Tab. IV.72: Vergleich der Befragungsergebnisse „überfachliche Kompetenzen" zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Zeiteffizienz" (1 = sehr schlechte Zeiteffizienz, 5 = sehr gute Zeiteffizienz)

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen		
		1	2	3
Moderne Sprachen	3304	3.67		
Musik	706		3.77	
Biologie und Chemie	1733		3.77	
Philosophie/Pädagogik/ Psychologie	1296		3.78	
Bildnerisches Gestalten	1135		3.78	
Wirtschaft und Recht	3145		3.80	3.80
Alte Sprachen	656		3.82	3.82
Physik und Anwendungen der Mathematik	1418			3.83

d max = 0.27

Tab. IV.73: Vergleich der Befragungsergebnisse „überfachliche Kompetenzen" zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „vernetztes Denken" (1 = sehr wenig vernetztes Denken, 5 = sehr viel vernetztes Denken)

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen				
		1	2	3	4	5
Moderne Sprachen	3299	3.20				
Physik und Anwendungen der Mathematik	1418		3.27			
Wirtschaft und Recht	3145		3.29	3.29		
Biologie und Chemie	1733		3.30	3.30		
Bildnerisches Gestalten	1135		3.30	3.30		
Musik	706			3.35	3.35	
Alte Sprachen	655				3.40	3.40
Philosophie/Pädagogik/ Psychologie	1296					3.44

d max = 0.31

Tab. IV.74: Vergleich der Befragungsergebnisse „überfachliche Kompetenzen" zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Perspektivenübernahme" (1 = sehr gute Fremdperspektiven-Übernahme bei Konflikten, 4 = sehr schlechte Fremdperspektiven-Übernahme bei Konflikten)

Schwerpunktfach	N (gewichtet)	Signifikante Untergruppen			
		1	2	3	4
Physik und Anwendungen der Mathematik	1409	2.92			
Wirtschaft und Recht	3139	2.92			
Biologie und Chemie	1733		2.98		
Moderne Sprachen	3295		3.01	3.01	
Philosophie/Pädagogik/ Psychologie	1296			3.05	3.05
Musik	705			3.05	3.05
Bildnerisches Gestalten	1134				3.06
Alte Sprachen	656				3.07

d max = 0.25

6.4 Vergleich zwischen den Geschlechtern

Wie dies bei den meisten Untersuchungen der Fall ist, wurden die Testergebnisse, die Maturanoten und die Befragungsergebnisse auch auf Unterschiede zwischen den Geschlechtern analysiert.

6.4.1 Geschlechtervergleich in Erstsprache

In Tabelle IV.75 findet sich der Vergleich der Leistungsergebnisse für Erstsprache. Es ergaben sich zwar mehrheitlich signifikant bessere Ergebnisse für die Maturandinnen, aber die Unterschiede bei den Testergebnissen sind unbedeutend. Hingegen wurden Maturandinnen im Ausmass einer leichten Effektstärke an der schriftlichen Maturaprüfung besser benotet. In der Erfahrungsnote ist der Unterschied zugunsten der Maturandinnen noch etwas höher.

Tab. IV.75: Vergleich Testergebnisse und Maturanoten zwischen den Geschlechtern in Erstsprache

Leistungsergebnisse	Geschlecht	N (gewichtet)	Mittelwert	Standard-abweichung	Mittlere Differenz	Signifikanz	Effektstärke
TOT_SPRACHE	männlich	5539	498	90.21	–4	0.009	0.05
	weiblich	7822	502	82.96			
ALLG_LV	männlich	5539	499	85.08	–2	0.140	0.03
	weiblich	7822	501	75.63			
DETAIL_LV	männlich	5539	495	89.39	–9	0.000	0.11
	weiblich	7822	504	80.86			
WORTS	männlich	5539	503	76.18	4	0.003	0.05
	weiblich	7822	499	70.22			
GRAM	männlich	5539	497	78.07	–6	0.000	0.08
	weiblich	7822	503	71.06			
E_ErfNote	männlich	5573	4.49	.520	–0.20	0.000	0.38
	weiblich	7824	4.69	.498			
E_müPr	männlich	5563	4.60	.760	–0.10	0.000	0.13
	weiblich	7820	4.70	.750			
E_schrPr	männlich	5571	4.37	.750	–0.07	0.000	0.10
	weiblich	7822	4.44	.730			
E_MatNote	männlich	5569	4.55	.547	–0.14	0.000	0.25
	weiblich	7820	4.69	.532			

6.4.2 Geschlechtervergleich in Mathematik

Der Vergleich für die Testergebnisse und die Maturanoten ist in Tabelle IV.76 dargestellt. Maturanden haben in allen Mathematik-Bereichen besser abgeschnitten. Die Unterschiede sind bei den Tests mittelhoch, bei den Noten hingegen nur noch klein (schriftliche Maturaprüfung) bis unbedeutend (Erfahrungsnote und mündliche Maturaprüfung).

Tab. IV.76: Vergleich Testergebnisse und Maturanoten zwischen den Geschlechtern in Mathematik

Leistungsergebnisse	Geschlecht	N (gewichtet)	Mittelwert	Standard-abweichung	Mittlere Differenz	Signifikanz	Effektstärke
TOT_MATH	männlich	5666	523	96.26	49	.000	0.56
	weiblich	7858	474	81.76			
ANALYSIS	männlich	5666	523	94.84	47	.000	0.54
	weiblich	7858	476	77.11			
GEOMETRIE	männlich	5666	525	91.94	49	.000	0.58
	weiblich	7858	476	76.35			
STOCHASTIK	männlich	5666	519	88.20	38	.000	0.45
	weiblich	7858	481	81.77			
VERSCH	männlich	5666	515	86.13	35	.000	0.42
	weiblich	7858	480	78.20			
Ma_ErfNote	männlich	5573	4.42	.787	.08	.000	0.11
	weiblich	7808	4.34	.795			
Ma_müPr	männlich	5091	4.43	1.385	.10	.000	0.08
	weiblich	7394	4.33	1.007			
Ma_schrPr	männlich	5570	4.19	1.052	.27	.000	0.25
	weiblich	7811	3.92	1.032			
Ma_MatNote	männlich	5569	4.42	.830	.13	.000	0.16
	weiblich	7820	4.29	.819			

6.4.3 Geschlechtervergleich in Biologie

In Tabelle IV.77 findet sich der Vergleich für die Leistungsergebnisse in Biologie und bei den Noten auch für weitere Bereiche der Naturwissenschaften. Wieder haben die Maturanden in allen Teilbereichen des Biologietests besser abgeschnitten, wobei die Unterschiede gering sind. Trotzdem wurden die Maturandinnen in der Erfahrungsnote in Biologie besser bewertet als die Maturanden. Dass in der Maturanote in Naturwissenschaften keine Unterschiede mehr zwischen den Geschlechtern auszumachen sind, kann auf das umgekehrt bessere Abschneiden der Maturanden in der Erfahrungsnote Physik zurückgeführt werden. Maturandinnen kompensierten offenbar die schlechteren Noten in Physik mit besseren Noten in Biologie. Die Unterschiede bei der Erfahrungsnote in Chemie sind unbedeutend.

Tab. IV.77: Vergleich Testergebnisse zwischen den Geschlechtern in Biologie und Maturanoten in Naturwissenschaften

Leistungs-ergebnisse	Geschlecht	N (gewichtet)	Mittelwert	Standard-abweichung	Mittlere Differenz	Signifikanz	Effektstärke
TOT_BIO	männlich	5573	507	87.56	17	0.000	0.19
	weiblich	7868	490	81.42			
B_EVO	männlich	5573	507	81.54	16	0.000	0.19
	weiblich	7868	491	76.41			
B_STRU	männlich	5573	507	84.08	16	0.000	0.19
	weiblich	7868	491	77.60			
B_GEN	männlich	5573	506	83.49	14	0.000	0.17
	weiblich	7868	492	78.32			
B_INFO	männlich	5573	506	84.01	15	0.000	0.19
	weiblich	7868	491	78.74			
B_OEKO	männlich	5573	506	84.50	16	0.000	0.20
	weiblich	7868	490	78.90			
B_STOFF	männlich	5573	506	86.06	15	0.000	0.18
	weiblich	7868	491	79.46			
Bi_ErfNote	männlich	4460	4.63	0.540	–.09	0.000	0.18
	weiblich	6652	4.72	0.539			
Ch_ErfNote	männlich	4511	4.60	0.633	.05	0.000	0.08
	weiblich	6715	4.55	0.603			
Ph_ErfNote	männlich	4413	4.59	0.684	.16	0.000	0.23
	weiblich	6740	4.43	0.697			
NW_MatNote	männlich	5565	4.60	0.575	.01	0.147	0.03
	weiblich	7805	4.59	0.560			
MPhyS_MatPr	männlich	1062	4.41	0.887	–.15	0.014	0.18
	weiblich	263	4.56	0.780			

6.4.4 Geschlechtervergleich bei den überfachlichen Fähigkeiten

Die grössten geschlechtsspezifischen Unterschiede ergeben sich beim überfachlichen Fähigkeitstest (UFT) (siehe Tabelle IV.78). Sie bewegen sich durchwegs in einem mittleren Ausmass. Vermutlich schlägt hier am meisten durch, dass es am Gymnasium relativ zur gleichmässigen Verteilung aller gleichaltrigen Adoleszenten in der Schweiz mehr Maturandinnen als Maturanden gibt, die Fähigkeitsvoraussetzungen aber quotenbezogen nicht besser sind, beziehungsweise eine höhere Quote am oberen Ende der (gleichen) Fähigkeitspyramide mit durchschnittlich tieferen Fähigkeiten verknüpft ist.

Tab. IV.78: Vergleich Testergebnisse zwischen den Geschlechtern im UFT

Leistungs-ergebnisse	Geschlecht	N (gewichtet)	Mittelwert	Standard-abweichung	Mittlere Differenz	Signifikanz	Effektstärke
U_QUANT	männlich	3833	525	82.60	48	.000	0.59
	weiblich	5448	477	78.84			
U_DIAGR	männlich	3833	524	79.91	47	.000	0.60
	weiblich	5448	477	78.08			
U_TEXT	männlich	3833	521	82.47	41	.000	0.51
	weiblich	5448	480	78.19			
U_NAT	männlich	3833	521	82.04	41	.000	0.51
	weiblich	5448	480	76.81			

6.4.5 Vergleich ausgewählter Ergebnisse aus der Zusatzbefragung

In Tabelle IV.79 sind nur jene Ergebnisse aufgeführt, bei denen die Effektstärke der Unterschiede in den selbst eingeschätzten überfachlichen Kompetenzen mindestens 0.2 beträgt. Während Maturandinnen angeben, das Lernen und die Prüfungsvorbereitungen zeitlich besser planen zu können und bei Teamkonflikten eher die Perspektiven der anderen verstehen zu wollen, behaupten Maturanden von sich, effizienter zu lernen.

Tab. IV.79: Signifikante Unterschiede zwischen den Geschlechtern mit Effektstärke > .2 bei überfachlichen Kompetenzen

Kompetenz-bereiche	Geschlecht	N (gewichtet)	Mittelwert	Standard-abweichung	Mittlere Differenz	Signifikanz	Effektstärke
Zeitplanung	männlich	5540	2.41	.944	–.39	.000	0.42
	weiblich	7856	2.80	.939			
Effiziente Prüfvor-bereitung	männlich	5540	3.49	.697	–.29	.000	0.45
	weiblich	7856	3.78	.614			
Zeiteffizienz	männlich	5540	3.88	.648	.20	.000	0.32
	weiblich	7856	3.68	.631			
Perspektiven-übernahme	männlich	5523	2.90	.612	–.15	.000	0.27
	weiblich	7847	3.05	.544			

6.5 Vergleich zwischen unterschiedlichen Maturitätsquoten

Die Maturitätsquoten in der Schweiz unterscheiden sich beträchtlich. Im Maturajahrgang 2007 haben gemäss Bundesstatistik im Kanton Tessin 29.5% der jungen Erwachsenen einen gymnasialen Abschluss erworben, im Kanton Obwalden waren es lediglich 11.2% (Bfs, 2008b). Diese Unterschiede müssen nicht zwingend eine entsprechende Chancen-Ungleichheit widerspiegeln. Es ist aber von Interesse, ob an den Maturitätsprüfungen überall die gleichen Anforderungen gestellt werden. Im Rahmen von EVAMAR II lassen sich Teile der Kompetenzen der Maturandinnen und Maturanden aus Kantonen mit unterschiedlichen Maturitätsquoten miteinander vergleichen. Weil die Stichprobe nicht für jeden Kanton repräsentativ ist, wird nur eine Grobeinteilung der Quoten zu Gruppen vorgenommen. Überlagerungen mit anderen Effekten bleiben dabei offen, ebenso die Frage der Gleichheit von Anforderungen bei der Eingangsselektion der Gymnasien. Bei der Beurteilung der Chancengleichheit müsste nämlich auch die Eingangsselektion berücksichtigt werden, weil ein „Aufholen“ von Kompetenzdefiziten während der Gymnasialzeit durchaus möglich ist, so dass sich dann Eingangsunterschiede am Schluss nicht mehr zeigen, die Chancen-Ungleichheit aber bleibt. Wegen der sich überlagernden Effekte wird im Folgenden ein Vergleich zwischen Gruppen in einem oberen und einem unteren Bereich vorgenommen. Orientiert man sich an einer möglichst gleichmässigen Dreiteilung, ergibt sich für das untere Drittel eine Quotengrenze von <17.5%, für das mittlere eine solche von 17.5% bis 18.9% und für das obere Drittel eine ab 19.0%. In den Tabellen IV.80 bis IV.83 finden sich die Ergebnisse der Vergleiche nach Maturitätsquoten.

Tab. IV.80: Darstellung von Unterschieden zwischen Kantonsgruppen mit unterschiedlichen Maturitätsquoten, Erstsprache

Leistungsergebnisse	Maturitätsquote	N (gewichtet)	Mittelwert	Standard-abweichung	Mittlere Differenz	Signifikanz	Effektstärke
TOT_SPRACHE	unteres Drittel	5483	508	83.69	12	.000	0.14
	oberes Drittel	4490	496	90.77			
ALLG_LV	unteres Drittel	5483	506	77.42	9	.000	0.12
	oberes Drittel	4490	497	83.44			
DETAIL_LV	unteres Drittel	5483	510	81.30	17	.000	0.19
	oberes Drittel	4490	493	89.81			
WORTS	unteres Drittel	5483	503	72.72	–2	.132	0.03
	oberes Drittel	4490	505	75.93			
GRAM	unteres Drittel	5483	505	77.00	7	.000	0.10
	oberes Drittel	4490	498	70.64			
E_ErfNote	unteres Drittel	5849	4.67	.469	.16	.000	0.31
	oberes Drittel	4497	4.51	.559			
E_müPr	unteres Drittel	5849	4.72	.721	.15	.000	0.20
	oberes Drittel	4487	4.57	.789			
E_schrPr	unteres Drittel	5849	4.50	.685	.20	.000	0.28
	oberes Drittel	4493	4.30	.784			
S_MatPr	unteres Drittel	5856	4.61	.614	.16	.000	0.25
	oberes Drittel	4487	4.45	.666			
E_MatNote	unteres Drittel	5849	4.70	.499	.16	.000	0.30
	oberes Drittel	4493	4.54	.572			

Tab. IV.81: Darstellung von Unterschieden zwischen Kantonsgruppen mit unterschiedlichen Maturitätsquoten, Mathematik

Leistungsergebnisse	Maturitäts-quote	N (gewichtet)	Mittelwert	Standard-abweichung	Mittlere Differenz	Signifikanz	Effektstärke
TOT_MATH	unteres Drittel	5673	504	89.73	24	.000	0.26
	oberes Drittel	4404	480	92.71			
ANALYSIS	unteres Drittel	5673	501	88.31	16	.000	0.18
	oberes Drittel	4404	485	88.06			
GEOMETRIE	unteres Drittel	5673	502	85.95	13	.000	0.15
	oberes Drittel	4404	489	87.76			
STOCHASTIK	unteres Drittel	5673	502	87.19	17	.000	0.19
	oberes Drittel	4404	485	84.06			
VERSCH	unteres Drittel	5673	507	83.72	30	.000	0.35
	oberes Drittel	4404	477	82.92			
Ma_ErfNote	unteres Drittel	5843	4.42	.765	.15	.000	0.20
	oberes Drittel	4497	4.27	.800			
Ma_müPr	unteres Drittel	5066	4.39	1.016	–.01	.706	0.01
	oberes Drittel	4371	4.40	1.410			
Ma_schrPr	unteres Drittel	5840	4.07	1.048	.12	.000	0.11
	oberes Drittel	4492	3.95	1.045			
Ma_MatNote	unteres Drittel	5849	4.37	.823	.09	.000	0.11
	oberes Drittel	4493	4.28	.810			

Tab. IV.82: Darstellung von Unterschieden zwischen Kantonsgruppen mit unterschiedlichen Maturitätsquoten, Biologie und Naturwissenschaften

Leistungsergebnisse	Maturitäts-quote	N (gewichtet)	Mittelwert	Standard-abweichung	Mittlere Differenz	Signifikanz	Effektstärke
TOT_BIO	unteres Drittel	5564	517	76.85	50	.000	0.61
	oberes Drittel	4462	467	86.02			
B_EVO	unteres Drittel	5564	515	70.63	44	.000	0.58
	oberes Drittel	4462	471	81.35			
B_STRU	unteres Drittel	5564	516	72.77	46	.000	0.59
	oberes Drittel	4462	470	83.12			
B_GEN	unteres Drittel	5564	515	73.03	43	.000	0.55
	oberes Drittel	4462	472	83.32			
B_INFO	unteres Drittel	5564	516	72.78	47	.000	0.60
	oberes Drittel	4462	469	84.35			
B_OEKO	unteres Drittel	5564	515	73.45	46	.000	0.57
	oberes Drittel	4462	469	85.58			
B_STOFF	unteres Drittel	5564	516	73.49	47	.000	0.59
	oberes Drittel	4462	469	86.04			
Bi_ErfNote	unteres Drittel	5841	4.67	.528	–.08	.000	0.15
	oberes Drittel	2377	4.75	.556			
Ch_ErfNote	unteres Drittel	5839	4.55	.597	–.01	.632	0.01
	oberes Drittel	2492	4.56	.654			
Ph_ErfNote	unteres Drittel	5841	4.46	.689	–.03	.060	0.04
	oberes Drittel	2479	4.49	.718			
NW_MatNote	unteres Drittel	5825	4.62	.572	.08	.000	0.13
	oberes Drittel	4493	4.54	.561			

Tab. IV.83: Darstellung von Unterschieden zwischen Kantonsgruppen mit unterschiedlichen Maturitätsquoten, UFT

Leistungsergebnisse	Maturitäts-quote	N (gewichtet)	Mittelwert	Standardab-weichung	Mittlere Differenz	Signifikanz	Effektstärke
U_QUANT	unteres Drittel	5162	503	82.93	15	.000	0.19
	oberes Drittel	832	488	78.69			
U_DIAGR	unteres Drittel	5162	503	81.78	17	.000	0.22
	oberes Drittel	832	486	75.28			
U_TEXT	unteres Drittel	5162	504	81.65	21	.000	0.26
	oberes Drittel	832	483	76.27			
U_NAT	unteres Drittel	5162	504	80.39	19	.000	0.25
	oberes Drittel	832	485	74.35			

Die Maturandinnen und Maturanden aus der Gruppe mit den tieferen Maturitätsquoten erzielen in beinahe allen Testbereichen die leicht besseren Resultate. Die grössten Unterschiede lassen sich bei den Ergebnissen in Biologie finden (mindestens mittlerer Effekt). Bei den meisten Maturanoten fallen die Unterschiede in der gleichen Richtung aus wie bei den Testergebnissen, aber in geringerem Ausmass. Ausnahme davon sind die Erstsprachenoten, deren Differenzen die Unterschiede in den Testergebnissen in Erstsprache sogar übersteigen. Auch wenn man die Grenzen wie Oelkers (2008, S. 64) mit <15%, 15% bis 20% und >20% für die drei Quotenbereiche setzt, fallen die Vergleiche in der gleichen Richtung aus. Es kann somit festgehalten werden, dass höhere Maturitäts-

quoten mehrheitlich mit schlechteren Ergebnissen der in EVAMAR II erfassten Teile der Maturitätsausbildung verknüpft sind und sich damit die Frage der Chancengleichheit stellt.

6.6 Vergleiche nach Strukturelement Lang- oder Kurzzeitgymnasium

Ein Teil der Schweizer Gymnasien schliesst direkt an die Primarstufe an und integriert die Sekundarstufe I in ein Langgymnasium. Es drängte sich deshalb auf zu untersuchen, ob Unterschiede in den Resultaten zu den Kurzgymnasien bestehen. Die Vergleiche finden sich in den Tabellen IV.84 bis IV.87.

Tab. IV.84: Unterschiede zwischen Lang- und Kurzzeitgymnasien: Erstsprache

Leistungsergebnisse	Dauer	N (gewichtet)	Mittelwert	Standard-abweichung	Mittlere Differenz	Signifikanz	Effektstärke
TOT_SPRACHE	Kurzzeit_G	11365	500	86.62	–5	.008	0.07
	Langzeit_G	1996	505	82.68			
ALLG_LV	Kurzzeit_G	11365	499	79.99	–9	.000	0.12
	Langzeit_G	1996	508	77.50			
DETAIL_LV	Kurzzeit_G	11365	499	85.18	–12	.000	0.14
	Langzeit_G	1996	511	80.75			
WORTS	Kurzzeit_G	11365	501	72.78	0	.998	0.00
	Langzeit_G	1996	501	72.78			
GRAM	Kurzzeit_G	11365	498	73.11	–16	.000	0.22
	Langzeit_G	1996	514	78.03			
E_ErfNote	Kurzzeit_G	11264	4.59	.523	–.10	.000	0.19
	Langzeit_G	2133	4.69	.475			
E_müPr	Kurzzeit_G	11254	4.66	.757	–.02	.212	0.03
	Langzeit_G	2129	4.68	.746			
E_schrPr	Kurzzeit_G	11260	4.39	.745	–.14	.000	0.19
	Langzeit_G	2133	4.53	.695			
E_MatNote	Kurzzeit_G	11260	4.62	.546	–.07	.000	0.14
	Langzeit_G	2129	4.69	.519			

Tab. IV.85: Unterschiede zwischen Lang- und Kurzzeitgymnasien: Mathematik

Leistungsergebnisse	Dauer	N (gewichtet)	Mittelwert	Standard-abweichung	Mittlere Differenz	Signifikanz	Effektstärke
TOT_MATH	Kurzzeit_G	11432	493	92.61	–10	.000	0.11
	Langzeit_G	2092	503	84.59			
ANALYSIS	Kurzzeit_G	11432	494	88.33	–9	.000	0.11
	Langzeit_G	2092	503	86.46			
GEOMETRIE	Kurzzeit_G	11432	496	87.62	–3	.109	0.04
	Langzeit_G	2092	499	81.01			
STOCHASTIK	Kurzzeit_G	11432	495	87.13	–10	.000	0.11
	Langzeit_G	2092	505	83.37			
VERSCH	Kurzzeit_G	11432	493	84.31	–13	.000	0.16
	Langzeit_G	2092	506	77.47			
Ma_ErfNote	Kurzzeit_G	11258	4.37	.791	–.03	.114	0.04
	Langzeit_G	2123	4.40	.802			
Ma_müPr	Kurzzeit_G	10356	4.37	1.208	–.01	.527	0.02
	Langzeit_G	2129	4.38	1.014			
Ma_schrPr	Kurzzeit_G	11249	4.01	1.051	–.13	.000	0.12
	Langzeit_G	2132	4.14	1.025			
Ma_MatNote	Kurzzeit_G	11260	4.33	.825	–.05	.012	0.06
	Langzeit_G	2129	4.38	.831			

Tab. IV.86: Unterschiede zwischen Lang- und Kurzzeitgymnasien: Biologie und Naturwissenschaften

Leistungsergebnisse	Dauer	N (gewichtet)	Mittelwert	Standard-abweichung	Mittlere Differenz	Signifikanz	Effektstärke
TOT_BIO	Kurzzeit_G	11433	493	84.51	–27	.000	0.32
	Langzeit_G	2008	520	80.18			
B_EVO	Kurzzeit_G	11433	494	78.86	–24	.000	0.31
	Langzeit_G	2008	518	76.34			
B_STRU	Kurzzeit_G	11433	494	80.80	–24	.000	0.30
	Langzeit_G	2008	518	77.19			
B_GEN	Kurzzeit_G	11433	494	80.88	–24	.000	0.30
	Langzeit_G	2008	518	77.21			
B_INFO	Kurzzeit_G	11433	493	81.64	–25	.000	0.31
	Langzeit_G	2008	518	76.15			
B_OEKO	Kurzzeit_G	11433	493	81.80	–25	.000	0.31
	Langzeit_G	2008	518	77.55			
B_STOFF	Kurzzeit_G	11433	493	82.96	–25	.000	0.31
	Langzeit_G	2008	518	76.90			
Bi_ErfNote	Kurzzeit_G	8979	4.70	.539	.09	.000	0.17
	Langzeit_G	2133	4.61	.546			
Ch_ErfNote	Kurzzeit_G	9093	4.56	.612	–.05	.002	0.08
	Langzeit_G	2133	4.61	.628			
Ph_ErfNote	Kurzzeit_G	9020	4.49	.701	–.05	.003	0.07
	Langzeit_G	2133	4.54	.673			
NW_MatNote	Kurzzeit_G	11236	4.59	.565	–.03	.077	0.04
	Langzeit_G	2133	4.62	.574			

Tab. IV.87: Unterschiede zwischen Lang- und Kurzzeitgymnasien: UFT

Leistungsergebnisse	Dauer	N (gewichtet)	Mittelwert	Standard-abweichung	Mittlere Differenz	Signifikanz	Effektstärke
U_QUANT	Kurzzeit_G	7196	495	82.14	–8	.000	0.09
	Langzeit_G	2085	503	88.99			
U_DIAGR	Kurzzeit_G	7196	495	81.38	–7	.000	0.09
	Langzeit_G	2085	502	84.73			
U_TEXT	Kurzzeit_G	7196	496	81.07	–6	.001	0.08
	Langzeit_G	2085	502	86.87			
U_NAT	Kurzzeit_G	7196	495	79.91	–7	.002	0.08
	Langzeit_G	2085	502	86.53			

Der Vergleich zeigt, dass in fast allen Testbereichen und auch bei den Maturanoten die Langzeitgymnasien zwar signifikant besser abschneiden, die Effektstärken aber sehr gering sind und nur in wenigen Bereichen das Mass von .2 (geringer Effekt) überschreiten. Sie finden sich durchwegs im Biologietest (nicht aber in den Biologienoten) und im Teil „Grammatik“ des Erstsprachtests. Vermutlich wird bei Letzterer am Langzeitgymnasium ein solideres Fundament gelegt als in der Sekundarschule. Ausserdem findet möglicherweise in Biologie am ehesten ein systematischer, über die ganze Dauer des Gymnasiums geplanter Aufbau statt, der zu einer breiteren Wissensbasis in diesem Fach führt, wenn das Gymnasium bereits die Sekundarstufe I einschliesst. In Kurzzeitgymnasien ist es schwierig, im Klassenverband an die sehr unterschiedlichen, bereits auf der Sekundarstufe I erreichten individuellen Biologiekompetenzen anzuknüpfen. In Kurzzeitgymnasien müssen Biologielehrpersonen deshalb in der Regel wieder bei den absoluten Grundlagen beginnen. Diese Übertrittsverluste sind wahrscheinlich bei den Kernfächern wie Erstsprache und Mathematik geringer, da hier eine grössere Einheitlichkeit bei den auf der Sekundarstufe I zu erreichenden Kompetenzen herrscht und die Lehrpersonen nicht wieder „bei null“ anfangen müssen. Interessanterweise ist aber auf der anderen Seite die Erfahrungsnote in Biologie die einzige Note, die für die Kurzgymnasien vorteilhafter ausfällt (siehe Tabelle IV.86). Vermutlich zeigt sich in diesem widersprüchlichen Ergebnis am deutlichsten die teilweise Anpassung der Notengebung an die Leistungsfähigkeit der Klasse, wenn klare Anforderungsstandards fehlen.

6.7 Weitere Vergleiche

6.7.1 Anstrengungsbereitschaft und Testergebnisse

Im ergänzenden Zusatzfragebogen wurden die Maturandinnen und Maturanden nach der Sorgfalt der Testbearbeitung gefragt. Zur Aussage „Ich habe die Aufgaben in den Tests sorgfältig bearbeitet“ konnten sie auf einer vierstufigen Skala angeben, inwieweit das der Fall war: „trifft nicht zu“, „trifft eher nicht zu“, „trifft eher zu“ und „trifft völlig zu“. Werden die Testergebnisse nach diesen Gruppen verglichen, ergibt sich durchwegs, dass mit ansteigender Sorgfalt der Bearbeitung auch das Testergebnis besser wird. Dies ist in Tabelle IV.88 exemplarisch für Mathematik gezeigt.

Tab: IV.88: Vergleich der Ergebnisse in Mathematik nach Sorgfalt der Testbearbeitung

Testsorgfalt	N (gewichtet)	Signifikante Untergruppen		
		1	2	3
trifft nicht zu	553	477		
trifft eher nicht zu	2791	478		
trifft eher zu	7686		496	
trifft völlig zu	1905			522

d max = 0.48

Dieses deutliche Ergebnis ist auf den ersten Blick im Hinblick auf die Validität der EVAMAR-Tests nicht unproblematisch. Deshalb wurden auch die erhobenen Maturanoten für diese Gruppen miteinander verglichen. Sie zeigen in allen Testfächern dieselbe Reihenfolge in der Notenhöhe. Dies sei wieder exemplarisch in den Tabellen IV.89 bis IV.91 für die Maturanoten in Mathematik gezeigt. Es kann somit gefolgert werden, dass mangelnde Sorgfalt die Testergebnisse nicht oder nur unwesentlich verfälscht hat und mangelnde Sorgfalt bei der Testbearbeitung typisch ist für jene Maturandinnen und Maturanden, die auch im Schulalltag und bei den Maturitätsprüfungen die schlechteren Resultate erzielt haben.

Tab. IV.89: Vergleich der Erfahrungsnote in Mathematik nach Sorgfalt der Testbearbeitung

Testsorgfalt	N (gewichtet)	Signifikante Untergruppen		
		1	2	3
trifft nicht zu	570	4.14		
trifft eher nicht zu	2705	4.15		
trifft eher zu	7546		4.42	
trifft völlig zu	1867			4.60

d max = 0.60

Tab. IV.90: Vergleich der schriftlichen Maturaprüfung in Mathematik nach Sorgfalt der Testbearbeitung

Testsorgfalt	N (gewichtet)	Signifikante Untergruppen			
		1	2	3	4
trifft nicht zu	568	3.73			
trifft eher nicht zu	2705		3.83		
trifft eher zu	7557			4.09	
trifft völlig zu	1866				4.28

d max = 0.53

Tab. IV.91: Vergleich der Maturanoten in Mathematik nach Sorgfalt der Testbearbeitung

Testsorgfalt	N (gewichtet)	Signifikante Untergruppen		
		1	2	3
trifft nicht zu	568	4.12		
trifft eher nicht zu	2705	4.15		
trifft eher zu	7558		4.38	
trifft völlig zu	1866			4.58

d max = 0.57

6.7.2 Vergleich von mündlichen und schriftlichen Noten

An den eigentlichen Maturitätsprüfungen werden in einigen Fächern sowohl schriftliche wie auch mündliche Prüfungen durchgeführt. In diesem Zusammenhang ist die Frage interessant, ob die beiden Noten Verschiedenes messen. Für die Testfächer wurden deshalb die Zusammenhänge zweifach untersucht. Mit einem t-Test für gepaarte Stichproben wurden die Unterschiede auf Signifikanz geprüft, und es wurde die Korrelation zwischen mündlicher und schriftlicher Note an der Maturitätsprüfung errechnet. Die Ergebnisse sind aus Tabelle IV.92 ersichtlich.

Tab. IV.92: Mittelwertsvergleich von mündlichen und schriftlichen Noten

Noten	Mittelwert	N (gewichtet)	Mittelwertsdifferenz	Signifikanz	Korrelation
Ma_müPr	4.36	12474	.31	.000	.608
Ma_schrPr	4.05	12474			
E_müPr	4.66	13383	.25	.000	.436
E_schrPr	4.41	13383			
Bi_müPR	4.93	171	.45	.000	.489
Bi_schrPr	4.48	171			
MphyS_müPr	4.50	1259	.25	.000	.537
MphyS_schrPr	4.25	1259			
BiChS_müPr	4.57	1499	.41	.000	.504
Bi_ChS_schrPr	4.16	1499			
MEF_müPr	4.36	91	.02	.726	.808
MEF_schrPr	4.34	91			
BiEF_müPr	5.24	129	.50	.000	.355
BiEF_schrPr	4.74	129			

Die Korrelationen sind überwiegend mässig und nur beim Ergänzungsfach Mathematik hoch. Mündliche und schriftliche Prüfungen messen offenbar nur teilweise dasselbe. Die mündlichen Noten fallen zudem fast durchwegs besser aus als die schriftlichen. Einzig beim Ergänzungsfach Mathematik unterscheiden sie sich nicht signifikant.

7 Ausgewählte korrelative Zusammenhänge

Das im Projekt EVAMAR II verwendete Testinstrumentarium umfasst verschiedenartige Prädiktoren des Studienerfolgs. Das Hauptgewicht des Tests in Erstsprache liegt auf übergreifenden, für alle Studienrichtungen wesentlichen Sprachkompetenzen. Der Mathematiktest enthält vor allem Aufgaben zu Inhalten, die für eine breite Anzahl von Studienfächern und nicht speziell für das Studienfach Mathematik von Bedeutung sind. Der Biologietest hat den Charakter eines Wissenstests zu Fachinhalten, die vor allem für ein Biologie- oder Medizinstudium, aber auch teilweise für nicht-naturwissenschaftliche Studien wichtig sind. Der überfachliche Fähigkeitstest (UFT) lehnt sich an die Eignungsprüfung für das Medizinstudium an und misst allgemeinere kognitive Fähigkeiten. Es handelt sich somit um die Messung verschiedenartiger Aspekte. Besonders interessant sind dabei die Zusammenhänge zwischen den verschiedenen Leistungsergebnissen; sie werden im Folgenden untersucht (Abschnitt 7.1).

Mit Ausnahme des Medizinstudiums ist in der effektiven Zulassungspraxis der Schweiz das Maturitätszeugnis ausreichende Grundlage für die Zugangsberechtigung zu den Schweizer Universitäten. Deshalb sollen auch die Zusammenhänge zwischen den Leistungstests von EVAMAR II und den effektiven Maturanoten untersucht werden (Abschnitt 7.2).

7.1 Zusammenhänge zwischen den Tests

In Tabelle IV.93 sind die Korrelationen zwischen den Gesamtergebnissen der Tests aufgeführt. Für den UFT wurde dazu über die Rasch-Modellierung ebenfalls ein Gesamtwert geschätzt.

Tab. IV.93: Korrelative Zusammenhänge zwischen den Tests

Leistungsergebnisse	TOT_MATH	TOT_BIO	U_GESAMT
TOT_SPRACHE	.148(**)	.224(**)	.298(**)
N (gewichtet)	12964	13182	8740
TOT_MATH		.223(**)	.448(**)
N (gewichtet)		12994	9082
TOT_BIO			.362(**)
N (gewichtet)			8860

** = signifikant auf dem 1%-Niveau

Alle Korrelationen sind signifikant, in der Höhe aber höchstens mittel (UFT und Mathematiktest mit r = .448). Dieses Ergebnis deutet darauf hin, dass im Projekt EVAMAR II – wie geplant – unterschiedliche Aspekte der Studierfähigkeit erfasst wurden.

7.2 Zusammenhänge der Tests mit den Maturanoten

In den folgenden Tabellen IV.94 bis IV.96 sind die korrelativen Beziehungen zwischen den Tests und den zugehörigen Fächern aufgeführt.

Tab. IV.94: Korrelationen zwischen dem Test und den Maturanoten in Erstsprache

Leistungsergebnisse	E_ErfNote	E_müPr	E_schrPr	E_MatNote
TOT_SPRACHE	.255(**)	.193(**)	.240(**)	.272(**)
N (gewichtet)	12803	12789	12800	12796
E_ErfNote		.474(**)	.616(**)	.836(**)
N (gewichtet)		13383	13394	13390
E_müPr			.436(**)	.732(**)
N (gewichtet)			13383	13383
E_schrPr				.787(**)
N(gewichtet)				13390

** = signifikant auf dem 1%-Niveau

Tab. IV.95: Korrelationen zwischen dem Test und den Maturanoten in Mathematik

Leistungsergebnisse	Ma_ErfNote	Ma_müPr	Ma_schrPr	Ma_MatNote
TOT_MATH	.395(**)	.296(**)	.458(**)	.456(**)
N (gewichtet)	12938	12092	12938	12947
Ma_ErfNote		.539(**)	.752(**)	.911(**)
N (gewichtet)		12468	13365	13373
Ma_müPr			.608(**)	.698(**)
N (gewichtet)			12474	12485
Ma_schrPr				.870(**)
N (gewichtet)				13378

** = signifikant auf dem 1%-Niveau

Tab. IV.96: Korrelationen zwischen dem Test und den Maturanoten in Biologie

Leistungsergebnisse	Bi_ErfNote	NW_MatNote
TOT_BIO	.250(**)	.257(**)
N (gewichtet)	10610	12834
Bi_ErfNote		.714(**)
N (gewichtet)		11088

** = signifikant auf dem 1%-Niveau

Die Korrelationen zwischen den Tests und den Maturanoten sind durchwegs signifikant, gehen aber über ein mittleres Mass nicht hinaus (am höchsten bei Mathematik schriftlich; r = .458). Am tiefsten ist der Zusammenhang bei Erstsprache. Es gibt wohl zwei Ursachen für das Ausbleiben höherer Korrelationen: Erstens sind die

Noten nicht standardisiert. Zweitens stimmen die Testinhalte und die mit den Maturitätsprüfungen erfassten curricularen Inhalte nur teilweise miteinander überein. Dass die Übereinstimmung in Erstsprache am tiefsten und in Mathematik am höchsten ist, entspricht den Erwartungen aus folgendem Grund: Die im Erstsprachtest erfassten Sprachfähigkeiten sind nicht gleichzusetzen mit den tatsächlichen curricularen Inhalten im Deutsch- oder Französischunterricht und haben auch überfachlichen Charakter. An den Maturitätsprüfungen in Erstsprache wird weit mehr und über die Relevanz für die allgemeine Studierfähigkeit Hinausgehendes geprüft (siehe Kapitel V). Das erfasste Wissen und Können weicht im Vergleich zu Mathematik und Biologie am weitesten von den mit den Tests erfassten Kompetenzen ab.

In einer weiteren Tabelle IV.97 sind fachweise die korrelativen Beziehungen zwischen dem UFT und den Maturanoten aufgeführt. Der UFT korreliert mit den Maturanoten signifikant, aber durchwegs tiefer als die Fachtests.

Tab. IV.97: Korrelationen zwischen dem UFT und den Maturanoten in Erstsprache

	Erfahrungsnote	Matura mündliche Prüfung	Matura schriftliche Prüfung	Maturanote
U_GESAMT und Erstsprache	.101(**)	.122(**)	.198(**)	.155(**)
N (gewichtet)	8740	8735	8740	8735
U_GESAMT und Mathematik	.287(**)	.227(**)	.332(**)	.303(**)
N (gewichtet)	8724	7947	8729	8735
U_GESAMT und Biologie	.227(**)			
N (gewichtet)	8501			
U_GESAMT und Chemie	.277(**)			
N (gewichtet)	8607			
U_GESAMT und Physik	.269(**)			
N (gewichtet)	8580			
U_GESAMT und Naturwissenschaften				.289(**)
N (gewichtet)				8715

** = signifikant auf dem 1%-Niveau

7.3 Zusammenhänge der Maturanoten untereinander

Eine häufig gestellte Frage ist jene nach der Heterogenität der mit Noten gemessenen Leistungen der Maturandinnen und Maturanden. In Tabelle IV.98 sind die korrelativen Zusammenhänge zwischen Maturanoten in den Testfächern aufgeführt. Sie sind wieder signifikant und höchstens mittelgross (Naturwissenschaften und Mathematik; r = .582).

Tab. IV.98: Korrelationen der Maturanoten untereinander

	Ma_MatNote	Bi_ErfNote	NW_MatNote
E_MatNote	.203(**)	.287(**)	.271(**)
N (gewichtet)	13390	11108	13365
Ma_MatNote		.396(**)	.582(**)
N (gewichtet)		11108	13365

** = signifikant auf dem 1%-Niveau

8 Zusammenfassung

8.1 Grundlagen der Test- und Befragungsergebnisse

Die Test- und Befragungserhebungen bei einer Stichprobe von rund 3'800 Maturandinnen und Maturanden aus der Deutschschweiz und der Romandie stützen sich auf folgende Grundlagen: Analyse von Lehrmaterialien und ersten Zwischenprüfungen der 16 gemessen an den Studierendenzahlen grössten universitären Studienfächer der Schweiz, Befragung der Dozierenden der Lehrveranstaltungen dieser Fächer sowie Einbezug aktueller Theorie und Forschung zur Messung kognitiver Leistungen und zur Voraussage des Studienerfolgs.

Das Hauptgewicht des Tests in Erstsprache liegt auf übergreifenden, für alle Studienrichtungen wesentlichen Sprachkompetenzen.

Der Mathematiktest enthält vor allem Aufgaben zu Inhalten, die für eine breite Anzahl von Studienfächern und nicht speziell für das Studienfach Mathematik von Bedeutung sind.

Der Biologietest hat den Charakter eines Wissenstests zu Fachinhalten, die vor allem für ein Biologie- oder Medizinstudium, aber auch teilweise für nicht-naturwissenschaftliche Studien wichtig sind.

Der überfachliche Fähigkeitstest (UFT) lehnt sich an die Eignungsprüfung für das Medizinstudium an und misst allgemeinere kognitive Fähigkeiten. Er wurde nur ergänzend in der Deutschschweiz durchgeführt.

Das Testinstrumentarium umfasst somit verschiedenartige Prädiktoren des Studienerfolgs und kann deshalb als ausgewogen bezeichnet werden. Es berücksichtigt die zu Beginn dieses Kapitels (siehe Abschnitt 1.2.1) beschriebenen Aspekte der Studierfähigkeit „allgemeine kognitive Leistungsfähigkeit", „bereits erworbenes Wissen" und „Lesefähigkeit". Interpretiert man allgemeine Studierfähigkeit dahingehend, jedes beliebige Studium ohne Probleme aufnehmen zu können, haben die in den Testbereichen gemessenen Kompetenzen folgende Bedeutung: Erstsprachkompetenzen gehören zu den Grundlagen beinahe jedes Studienfachs. Mathematik ist wichtig für eine grosse Zahl von Fächern, in denen die mathematische Formalsprache verwendet wird. Biologisches Wissen braucht es nur für eine eingeschränkte Zahl von Studienfächern. Die mit dem UFT gemessenen Fähigkeiten korrelieren nachgewiesenermassen signifikant mit dem Studienerfolg im Medizinstudium und damit mit jenem in allen anderen Studienrichtungen, die gleiche oder ähnliche Fähigkeiten erfordern. Die Voranalysen haben im Weiteren ergeben, dass es allgemein für jedes Fachstudium von Vorteil ist, bereits zu Beginn über Wissen und Können aus dem entsprechenden Maturitätsfach zu verfügen, wenn auch in unterschiedlichem Umfang.

8.2 Gesamtergebnisse der Tests

Im Mittel wurden die Testaufgaben etwa zur Hälfte richtig gelöst; der Anteil der richtig gelösten Aufgaben und die geschätzten Personenfähigkeiten waren in Erstsprache etwas höher als in Mathematik und Biologie. In Biologie fielen die Ergebnisse für die Gruppe der Maturandinnen und Maturanden, die dieses Fach bereits ein halbes Jahr zuvor oder länger abgeschlossen hatten, klar schlechter aus, vermutlich infolge des Vergessenseffekts. Das deutet darauf hin, dass das an Maturitätsschulen erworbene Wissen nur für kurze Zeit vollumfänglich präsent ist, und dass sowohl ein möglichst später Abschluss des Fachs am Gymnasium wie auch eine rasche Aufnahme des Studiums vorteilhaft sind. Die Ergebnisse des UFT liegen nur leicht unter den Resultaten des jährlich durchgeführten Medizineignungstests. Das ist plausibel, weil bei den EVAMAR-Erhebungen auch all jene Personen teilgenommen haben, die sich ein Medizinstudium wegen mangelnder Fähigkeiten nicht zutrauen. Dieses Ergebnis ist zudem ein wichtiger Indikator für die hohe Anstrengungsbereitschaft, welche die Maturandinnen und Maturanden bei der Bearbeitung der EVAMAR-Tests offenbar an den Tag legten. Aufgrund der Nähe der Ergebnisse des UFT zu den Resultaten, wie sie im individualbiografisch hoch relevanten Kontext der Eignungsprüfung erreicht werden, braucht die Anstrengungsbereitschaft deshalb auch den Vergleich mit jener in den gewohnten notenwirksamen Testsituationen im Gymnasium keineswegs zu scheuen.

Sowohl zwischen Einzelpersonen als auch zwischen ganzen Klassen gibt es grosse Unterschiede in den Ergebnissen, die grösser ausfallen als bei den effektiv erteilten Maturanoten. Es lässt sich somit feststellen, dass erhebliche Anteile von Maturandinnen und Maturanden in den mit den Tests erfassten Bereichen über vermutlich ungenügende Kompetenzen in mindestens einem Testbereich verfügen. Diese Erkenntnis zeichnet sich aber auch schon bei einer Analyse der durch die Schulen selbst vergebenen Maturanoten ab, am ausgeprägtesten in Mathematik. 41.4% der Maturandinnen und Maturanden wurden im Jahr 2007 in der schriftlichen Maturaprüfung mit einer ungenügenden Note bewertet. Bei der Matura-Endnote in Mathematik, die auch die Erfahrungsnoten und die mündliche Prüfung berücksichtigt, waren es immer noch 24.4%. Zwar bleibt offen, ob es sich bei den für die ungenügenden Ergebnisse in den Maturitätsprüfungen ursächlichen Teile um Wissen und Können handelt, das für die allgemeine Studierfähigkeit unabdingbar ist. Im Extremfall könnten einerseits Maturandinnen und Maturanden mit ungenügenden Noten trotzdem in für die Studierfähigkeit relevanten Bereichen genügende Werte aufweisen. Andererseits könnte es auch Maturandinnen und Maturanden mit genügenden Noten geben, die in für die allgemeine Studierfähigkeit relevanten Bereichen des Fachs trotzdem über ungenügende Kompetenzen verfügen. Die teilweise schlechten Ergebnisse in den Tests hingegen betreffen durchwegs Teile, die für mindestens ein Studienfach relevant sind. Ab welcher Grenze bei den Leistungstests ein erfolgreiches Studium nicht mehr möglich ist, kann nicht deterministisch bestimmt werden. Dazu wären weitere, umfangreiche und langwierige empirische Untersuchungen notwendig. Die Vermutung aber, dass nicht alle Maturandinnen und Maturanden in der ganzen Breite über genügende Kompetenzen verfügen, um jedes beliebige Studium aufnehmen zu können, ohne zusätzlich Lücken füllen zu müssen, lässt sich – unter Einbezug der Maturanoten – auf gute Gründe stützen. Dies, obwohl sie durch die Verleihung des Maturaausweises als „allgemein studierfähig“ qualifiziert wurden. Diese Erkenntnis stimmt überein mit von Dozierenden subjektiv festgestellten Kompetenzlücken in den Bereichen Mathematik und Erstsprache (siehe Kapitel II, Abschnitt 3.3). Wegen des Kompensationssystems für ungenügende Noten an Maturaprüfungen ist dieses Ergebnis der Maturitätsbildung aber kaum vermeidbar.

8.3 Ergebnisse der Tests unter strukturellen Gesichtspunkten

Die Anzahl Jahre der gymnasialen Schuldauer an einem eigenen gymnasialen Schulort scheint sich auf die Testergebnisse auszuwirken. Eine Mindest-Dauer von mindestens vier Jahren an einem Gymnasium geht einher mit besseren Ergebnissen im Vergleich zu jener Ausbildungsstruktur, bei der das Erste von vier Jahren in einer „gymnasialen“ Klasse an der örtlichen Sekundarschule absolviert werden kann. Dies zeigt sich beim Mathematiktest, in leichterer Ausprägung auch bei Erstsprache und in der Deutschschweiz zusätzlich beim UFT.

Langzeitgymnasien schneiden in fast allen Testbereichen besser ab als Kurzzeitgymnasien, allerdings nur leicht. Einzig im Biologietest ergeben sich auch grössere bzw. beinahe mittlere Differenzen zugunsten der Langzeitgymnasien, obwohl bei Letzteren die Biologie-Erfahrungsnoten im Ausmass zwar leicht, aber signifikant schlechter ausgefallen sind. Dabei dürfte es sich aber eher um eine „Notenanomalie“ bzw. das Ergebnis einer klassenorientierten Benotung handeln.

Zwischen den verschiedenen Schwerpunktfachgruppen lassen sich zum Teil erhebliche Unterschiede feststellen. Die Spitzenergebnisse in Mathematik der Gruppe des Schwerpunktfachs (SPF) „Physik und Anwendungen der Mathematik“ (PAM) und in Biologie der Gruppe des SPF „Biologie und Chemie“ (BCH) waren dabei systembedingt zu erwarten. Der Spitzenwert für die Maturandinnen und Maturanden des SPF „Alte Sprachen“ im Vergleich zu jenen des SPF „Moderne Sprachen“, die beim Erstsprachtest lediglich eine mittlere Rangierung erreichten, überrascht jedoch teilweise. Auch für weitere Differenzen in einigen Kompetenzbereichen gibt es keine Systemerklärungen. Die Gruppe des SPF „Alte Sprachen“ hat unter dem Aspekt der Ausgeglichenheit bzw. Ausgewogenheit der Kompetenzen (im Sinne einer allgemeinen Studierfähigkeit) am besten abgeschnitten, die Maturandinnen und Maturanden der neuen Schwerpunktfächer „Musik“, „Bildnerisches Gestalten“, „Philosophie/Pädagogik/Psychologie“ (PPP) eher schlechter. Bei der Analyse der effektiv erteilten Maturanoten ergeben sich im Vergleich zu den Testergebnissen einige Rangverschiebungen, und vor allem erweisen sich die Unter-

schiede als geringer ausgeprägt. Es fällt aber auch auf, dass es SPF-Gruppen mit beinahe 50% („Moderne Sprachen“, „Bildnerisches Gestalten“, „Musik“) oder gar über 50% (PPP) ungenügenden Noten in der schriftlichen Mathematikprüfung gibt. Bei der Maturagesamtnote in Mathematik sind es für diese SPF-Gruppen (jetzt ohne die SPF-Gruppe „Musik“) immer noch rund 30% ungenügende Noten. Mit Abstand die beste Gruppe beim UFT war jene des SPF PAM. Die Gruppen der neuen Schwerpunktfächer PPP und „Bildnerisches Gestalten“ sowie in einem Bereich auch „Musik“ rangieren dagegen am Ende.

Die Unterschiede zwischen den Geschlechtern stimmen mit bisherigen Untersuchungen gut überein (z. B. Halpern & LaMay, 2000; Köller & Klieme, 2000; Stanat & Kunter, 2001; Dresel, Stöger & Ziegler, 2006). Maturandinnen haben eher in Erstsprache, Maturanden klar in Mathematik und im naturwissenschaftlich ausgerichteten UFT besser abgeschnitten. Allerdings weist die PISA-Studie 2003 nur noch für die Lesekompetenz deutlich schwächere Leistungen der Jungen in allen beteiligten Staaten nach (d max = 0.46); bei den Fähigkeiten in Naturwissenschaften und Problemlösen besteht kein signifikanter Geschlechtsunterschied mehr, und der Unterschied in der Mathematik taucht nur noch schwach (d max = 0.19) zugunsten der Jungen auf (Zimmer, 2006). Auch in der PISA-Studie 2006 wurden in der globalen Kompetenz Naturwissenschaften und Problemlösen keine signifikanten Geschlechtsunterschiede festgestellt, währenddessen der Unterschied bei den Lesefähigkeiten wie schon in PISA 2003 weiterhin stark und in Mathematik weiterhin schwach ausgeprägt ist (OECD, 2007). Das in EVAMAR II im Vergleich zu den PISA-Studien relativ bessere Abschneiden der männlichen Probanden lässt sich damit erklären, dass es am Gymnasium relativ zur gleichmässigen Verteilung aller gleichaltrigen Adoleszenten in der Schweiz mehr Maturandinnen als Maturanden gibt und eine höhere Quote am oberen Ende der Fähigkeitspyramide mit im Durchschnitt relativ tieferen Fähigkeiten verknüpft ist.

Zieht man einen Vergleich zwischen den Kantonen, die sich hinsichtlich der Maturitätsquoten im oberen Drittel befinden, und jenen im unteren Drittel, so zeigen sich leichte bis zwischen mittel und gross liegende Unterschiede. Die positiven Differenzen zugunsten der Kantone mit den niedrigsten Maturitätsquoten stimmen nur teilweise mit entsprechenden Unterschieden bei den Maturanoten überein, die aber grossmehrheitlich bei dieser Gruppe im Mittel ebenfalls besser ausfallen. Es stellt sich die Frage der Chancengleichheit beim Erwerb der Berechtigung zum Hochschulzugang.

8.4 Notengebung

Die Notengebung stimmt nur teilweise mit den Testergebnissen überein. Eine erste Erklärung dafür ist, dass mit den Tests teilweise andere Kompetenzen gemessen wurden als mit den Maturaprüfungen. Diese Unterschiede sind durch den beschränkten Fokus der Untersuchung bedingt und haben nichts mit mangelnder Zielerreichung der Gymnasien zu tun. Eine zweite Deutung der Befunde setzt beim Verfahren der Notengebung an. Einerseits kann eine teilweise Anpassung der Notengebung an die Leistungsfähigkeit der Klasse (Sozialnormorientierung) vermutet werden, was im Vergleich zur Orientierung an klassenübergreifenden Gütestandards (Kriteriumsorientierung) zu verschiedenen Notenmassstäben führt. Dennoch lassen sich andererseits auch bei den Notenvergleichen strukturelle Übereinstimmungen mit den Testergebnissen feststellen, was annehmen lässt, dass sich die Lehrerinnen und Lehrer bei der Notengebung zu einem beträchtlichen Teil auch an „objektivierten Gütestandards“ orientieren. Im Weiteren sind bei der Notengebung unerklärliche „Anomalien“ zu finden, welche wiederum eine nur teilweise Kriteriumsorientierung annehmen lassen. Dazu gehört z. B. die Erfahrungsnote in Biologie, die in den Schwerpunktgruppen „Alte Sprachen“ und „Bildnerisches Gestalten“ höher liegt als in der Gruppe „Biologie und Chemie“. Auffällig ist schliesslich, dass schriftliche Maturaprüfungen vor allem in Mathematik, aber auch in Erstsprache zu einem hohen Anteil ungenügender Ergebnisse führen, diese aber jeweils durch die Note der mündlichen Prüfungen und die Erfahrungsnoten „aufgebessert“ werden.

8.5 Eigene Wahrnehmung überfachlicher Kompetenzen

In Ergänzung zu den Studien von Notter & Arnold (2003 und 2006) wurden die Maturandinnen und Maturanden nach ihrer Wahrnehmung einiger ihrer überfachlichen Kompetenzen befragt. Dabei zeigten sich bei überwiegend zufriedenstellenden Werten auch von den Schülerinnen und Schülern selbst wahrgenommene Defizite, insbesondere bei den Bereichen „aktive Zeitplanung“ und „Konzentrationsfähigkeit“.

Auch bei Notter & Arnold (2006) rangiert die Kompetenz der „Zeiteinteilung“ aus Sicht der Studierenden am unteren Ende ihrer Kompetenzen (S. 31). Dabei messen sowohl die in dieser Studie befragten Studierenden (S. 52) als auch die bei EVAMAR II im Rahmen des Teilprojekts A4 befragten Dozierenden dieser Fähigkeit eine hohe Bedeutung bei (vgl. Kapitel II). Zudem beklagten ca. 15% der Dozierenden, die diese Fähigkeit als wichtig einstuften, bei ihren Studierenden entsprechende Mängel, ein Befund, der mit der Einschätzung der Maturandinnen und Maturanden korrespondiert.

Bezüglich der aus Sicht der Maturandinnen und Maturanden defizitären Konzentrationsfähigkeit liegen keine Vergleichswerte aus der Studierendenbefragung vor. Diese Kompetenz wurde erst in den Fragekatalog des Zusatzfragebogens aufgenommen, nachdem Dozierende sie in offener Nennung auf die Frage nach weiteren – das heisst nach noch nicht bereits von Notter & Arnold (2006) aufgeführten – Kompetenzen angeführt hatten (Teilprojekt A4). Bei den ebenfalls aufgrund dieser offenen Nennungen aufgenommenen Kompetenzen des vernetzten Denkens, der Ausdauer sowie der Perspektivenübernahme und der Diskussionsfähigkeit stellen sich die Maturandinnen und Maturanden im Durchschnitt ein gutes Zeugnis aus. Die letztgenannten beiden Kategorien ähneln dabei jener der „Sozialkompetenz“ bei Notter & Arnold (2006), die von den Studierenden in deren Untersuchung als ähnlich gut ausgeprägt eingeschätzt wurden. Auch die in Teilprojekt A4 befragten Dozierenden geben bezüglich dieser beiden für das Gelingen von Gruppenarbeit wichtigen Kompetenzen eine ähnliche Bewertung ab: So sahen ca. 97% hier kein Manko bei den Studierenden bezüglich der von ihnen als wichtig eingestuften Kompetenz „im Team arbeiten“. Im Vergleich dazu beurteilten die Dozierenden unserer Befragung die Kompetenzen der Zeiteffizienz und der effizienten Prüfungsvorbereitung ihrer Studierenden in den ersten Semestern als etwas defizitärer, jedoch ist der Anteil von ca. 8% letztlich nicht sehr hoch. Damit stimmen sie mit der Einschätzung der Maturandinnen und Maturanden überein, die diesbezüglich im Durchschnitt eher wenig Mühe bekunden.

In Bezug auf die zu Beginn dieses Kapitels (siehe Abschnitt 1.2.1) aufgeführten Prädiktoren der Persönlichkeit für die akademische Leistungsfähigkeit lässt sich festhalten, dass zumindest aus der Selbstwahrnehmung der Maturandinnen und Maturanden heraus eher Optimismus herrschen kann. Als in diesem Sinne für die akademische Leistungsfähigkeit förderliche Persönlichkeitsmerkmale wurden in EVAMAR II mit den Items zur Ausdauer Aspekte des Leistungsstrebens und mit Items innerhalb der Skala der effizienten Prüfungsvorbereitung solche der Selbstdisziplin erhoben. Während die recht hohen Werte bei diesen Items daher eine gute Voraussetzung für die Studierfähigkeit signalisieren, wäre eine Verbesserung beim Indikator der Konzentrationsfähigkeit sicherlich wünschenswert, da auch diese zu einer guten Selbstdisziplin zu zählen ist.

Der bei EVAMAR II erhobene Indikator des „vernetzten Denkens“ schliesslich stellt Elemente des tiefenverarbeitenden Verarbeitens von Informationen, des für die akademische Leistungsfähigkeit ebenfalls förderlichen Deep Processings als einem motivationalen Faktor dar. Wenngleich das vernetzte Denken aus Sicht der Schülerinnen und Schülern nicht so stark ausgeprägt ist wie etwa die effiziente Prüfungsvorbereitung, attestieren sich die Maturandinnen und Maturanden doch, „manchmal bis oft“ Verbindungen zwischen verschiedenen Lerninhalten zu ziehen. Sie deuten damit durchaus auch im Sinne des für akademische Leistungsfähigkeit weiterhin förderlichen Leistungsziels „Mastery Goals“ darauf hin, dass sie das Bedürfnis haben, den Lehrstoff wirklich zu verstehen und nicht nur auswendig zu lernen oder oberflächlich zu reüssieren.

V Projektteil D1: Analyse schriftlicher Maturaprüfungen

1 Allgemeines und Übersicht

1.1 Zielsetzungen und Auswahl der Fächer

Im Teilprojekt D1 von EVAMAR II sollten die schriftlichen Maturitätsprüfungen aller Kantone qualitativ untersucht werden im Hinblick darauf, welche fachinhaltlichen und welche kognitiven Anforderungen sie stellen und inwieweit es eine Übereinstimmung gibt mit den in den Teilprojekten A und B ermittelten Komponenten der Studierfähigkeit sowie mit den in Teilprojekt C durchgeführten Tests. Diese Analyse erfolgte zunächst für die Testfächer Erstsprache, Mathematik und Biologie; sie könnte in einem Folgeprojekt auf weitere Fächer ausgedehnt werden.

1.2 Allgemeines zu den Analyse-Instrumenten

Kognitive Anforderungen von Aufgaben können grundsätzlich mittels verschiedener Modelle untersucht werden. Im Projekt EVAMAR II stehen zwei Modelltypen im Vordergrund: Das bereits in Teilprojekt A verwendete allgemeine Modell von Anderson et al. (2001) und die in den Teilprojekten B und C übernommenen und weiterentwickelten fachspezifischen Modelle.

Für die Analyse der kognitiven Anforderungen drängte sich das Modell von Bloom et al. (1956 und 1974) bzw. die im Sinne einer Anpassung an neuere Erkenntnisse der Kognitionspsychologie sorgfältig überarbeitete Version von Anderson et al. (2001) (siehe Tabelle V.1) aus drei Gründen auf: Erstens erleichtert es vergleichende Aussagen über alle Fächer. Zweitens handelt es sich um eines der bestbewährten und meistverbreiteten pädagogisch-didaktischen Modelle der letzten Jahrzehnte[29], welchem gerade in der aktuellen Diskussion über Kompetenzmodelle wieder verstärkt Bedeutung zukommt (siehe z. B. Klieme & Leutner, 2006, S. 8). Es stellt den mit Abstand am besten etablierten und ausgereiften Analyseraster für kognitive Anforderungen schulischer Leistungen dar. Drittens werden fachspezifische Modelle trotzdem noch beigezogen, nämlich bei der fachinhaltlichen Analyse. Die Eignung des allgemeinen Modells für kognitive Prozesse wurde für jedes untersuchte Fach nochmals separat geprüft.

[29] Das ursprüngliche Modell wurde bislang in 22 Sprachen übersetzt.

Tab. V.1: Übersicht über die Haupt- und Unterkategorien kognitiver Prozesse nach Anderson et al. (2001) und die im Rahmen von EVAMAR II adaptierten Definitionen der Kategorien (verdeutlicht zunächst nur für das Fach Biologie).

Hauptkategorie und Teilkategorien	Kognitive Prozesse *[unmittelbar übersetzt aus Anderson et al. (2001), p. 63–92]*	Alternative Bezeichnungen *[unmittelbar übersetzt aus Anderson et al. (2001), p. 63–92]*	In EVAMAR II adaptierte Definitionen (zunächst für das Fach Biologie) *[angelehnt an Anderson & Krathwohl (2001), p. 63–92]*
1	Erinnern relevantes Wissen aus dem Langzeitgedächtnis abrufen		
1.1	wieder erkennen	identifizieren	Wissen, das mit dem vorgegebenen Material konsistent ist, im Langzeitgedächtnis auffinden
1.2	abrufen	erinnern	relevantes Wissen aus dem Langzeitgedächtnis abrufen
2	Verstehen aus Lehrinhalten (mündlich, schriftlich und grafisch) Bedeutung gewinnen		
2.1	interpretieren	klären, verdeutlichen, umschreiben, umformulieren, anders darstellen, übersetzen, übertragen, umarbeiten	vorgegebenes Material von einer Darstellungsform in eine andere bringen
2.2	Beispiele machen, veranschaulichen, beispielhaft zeigen	illustrieren, veranschaulichen, bildlich darstellen	ein bestimmtes Beispiel oder eine bestimmte Abbildung finden, welche(s) ein Konzept oder ein Prinzip veranschaulicht
2.3	klassifizieren	kategorisieren, subsumieren	entscheiden, welche Instanzen zu welchen Kategorien (Konzepten oder Prinzipien) gehören
2.4	zusammenfassen	abstrahieren, verallgemeinern	• zu vorgegebenem Material ein Hauptthema finden oder die wichtigsten Punkte zusammenfassen (mit eigener Integrationsleistung) • aus dem vorgelegten Material die Quintessenz, die Pointe, das „Worum-es-geht" herausfiltern und zusammenfassen (mit eigener Integrationsleistung)
2.5	folgern	schliessen, ableiten, extrapolieren, interpolieren, voraussagen, hochrechnen	aus dem vorgegebenen Material logische Schlüsse ziehen, im Sinne von: geg.: ein Satz von Beispielen oder Instanzen, oder geg.: Analogie in der Form A ist zu B wie C zu D ges.: durch Vergleichen Beziehungen, Konzepte oder Prinzipien zwischen den Beispielen oder Instanzen erkennen, also • Muster erkennen • Bildungsregel ableiten→ extrapolieren • Analogon D finden • nicht dazu passendes Element finden
2.6	vergleichen	kontrastieren, gegenüberstellen, zuordnen, Ähnlichkeiten suchen	• Übereinstimmungen und Unterschiede zwischen Ideen, Objekten u.Ä. erkennen • Ideen, Objekte u.Ä. innerhalb derselben Hierarchie-Ebene einander zuordnen (Mapping)

Hauptkategorie und Teilkategorien	Kognitive Prozesse *[unmittelbar übersetzt aus Anderson et al. (2001), p. 63–92]*	Alternative Bezeichnungen *[unmittelbar übersetzt aus Anderson et al. (2001), p. 63–92]*	In EVAMAR II adaptierte Definitionen (zunächst für das Fach Biologie) *[angelehnt an Anderson & Krathwohl (2001), p. 63–92]*
2.7	erklären	Modelle und Mechanismen konstruieren	ein zuvor erlerntes Ursache-Wirkungs-Modell auf bisher unbekannte Tatsachen anwenden Anwenden heisst: • geg. Wirkung(en): das Modell zum Schliessen auf Ursachen verwenden • geg. Ursache(n): das Modell zum Schliessen auf Wirkungen verwenden
3	Anwenden in einer gegebenen Situation eine Prozedur ausführen oder einsetzen		
3.1	ausführen	–	gemeinsame Kategorie 3.1/3.2 eine Prozedur (mechanisch einsetzbare und einübbare Schritte) auf ein mehr oder weniger bekanntes Problem anwenden
3.2	implementieren	anwenden, von etwas Gebrauch machen	
4	Analysieren vorgegebenes Material in seine Bestandteile zerlegen und bestimmen, wie die Teile untereinander in Beziehung stehen und wie die Teile mit der Gesamtstruktur oder einem Gesamtzweck zusammenhängen		
4.1	differenzieren	unterscheiden, auseinanderhalten, fokussieren, auswählen	relevante und irrelevante Teile innerhalb des vorgegebenen Materials unterscheiden (ohne eigene Integrationsleistung)
4.2	organisieren	Kohärenz finden, integrieren, grob skizzieren, analysieren, strukturieren	• bestimmen, wie Elemente innerhalb einer Struktur zusammenpassen oder zusammenspielen • vorgelegtes Material strukturieren (Zwischentitel setzen, Tabelle konstruieren, Box-Pfeil-Diagramm oder Mindmap erstellen usw.), also das Material auch in eine andere Form bringen
4.3	auf etwas zurückführen	dekonstruieren	Blickwinkel, Wahrnehmungsverzerrungen, Wertvorstellungen oder Absichten eines Autors von vorgegebenen Material erkennen
5	Bewerten auf Kriterien und Standards abgestützt ein Urteil abgeben		
5.1	überprüfen	in ein Zusammenspiel bringen, Fehler detektieren, überprüfen, testen	Inkonsistenzen oder Fehler innerhalb eines Prozesses oder Produkts erkennen; entscheiden, ob ein Prozess oder Produkt (in sich selber) konsistent ist; die Effektivität einer Prozedur (so wie sie angewandt wird) bestimmen
5.2	kritisch untersuchen	beurteilen	erkennen, wo ein Prozess oder ein Produkt, gemessen an externen Kriterien, inkonsistent ist; bestimmen, ob ein Produkt externe Konsistenz besitzt; erkennen, ob eine Prozedur bei einem gegebenen Problem angemessen ist
6	Erschaffen Elemente zusammensetzen, um ein kohärentes funktionales Ganzes zu formen; Elemente auf neuartige Weise strukturieren		
6.1	hervorbringen, entwickeln	hypothetisieren	sich Hypothesen ausdenken
6.2	planen	gestalten	eine Vorgehensweise entwickeln, um eine Aufgabe zu lösen
6.3	produzieren	konstruieren	ein Produkt neu erfinden

Die Aussagekraft der mittels dieses Analyserasters gewonnenen Ergebnisse hat Grenzen. Die fast unvermeidbare Unsicherheit, die bei der Kategorisierung von Prüfungsaufgaben auftritt, wurde in der pädagogischen Forschung bereits früh erkannt. Cazden (1986) fasst folgendermassen zusammen: „Thinking about questions in terms of some scale of cognitive difficulty is probably still heuristically useful for teachers, but inherently imprecise for research“ (p. 453). Die wichtigste Einschränkung ist die folgende: Das Anspruchsniveau einer Aufgabe ist immer auch abhängig vom vorausgegangenen Unterricht. Eine auf den ersten Blick taxonomisch anspruchsvolle Aufgabe kann durchaus auch als reine Wissenswiedergabe eingestuft werden, wenn deren Lösung in der Vorbereitung bereits (auswendig) gelernt werden konnte. „Teaching to the test“ bzw. ein „testing to the teaching“ kann bei „hausgemachten“ Maturaaufgaben gezielt vorgenommen werden. Ein Vergleich mit dem vorangehenden Unterricht war aber im Projekt EVAMAR II aus Ressourcengründen nicht möglich und im Auftragskonzept auch nicht vorgesehen. Auch Carlsen (1991) weist darauf hin, dass es forschungsökonomisch nicht denkbar ist, den gesamten Unterrichtsverlauf bis zur Prüfung (lückenlos) zu dokumentieren. Eine scheinbar attraktive Lösung des Problems ist die direkte Befragung der Lehrpersonen, welche die Prüfungsaufgaben verfasst haben. Allerdings zeigt sich, dass viele Lehrpersonen ungeübt sind in der Anwendung des Schemas nach Anderson et al. (2001) und überdies ihre Aufgaben ungern in die Kategorie 1.2 einstufen, da diese im Verdacht der Einfallslosigkeit und des unnötigen Belastens der Maturandinnen und Maturanden mit Detailwissen steht. Zudem wären zur Sicherstellung einer ausreichenden Reliabilität der Einstufungsentscheide entweder eine intensive vorgängige Schulung der Lehrpersonen in der Handhabung des Kategoriensystems oder – zu jeder einzelnen Aufgabe – aufwändige Einzelgespräche zwischen der Lehrperson und den Autoren von EVAMAR II nötig gewesen. Eine kürzlich in der Zeitschrift „Science“ veröffentlichte Studie stützt die Zuverlässigkeit der Kategorisierung ebenfalls ausschliesslich auf Codierer ab, welche den vorangegangenen Unterricht nicht kannten (Zheng et al., 2008). Die Autoren nehmen ebenso an, dass die oben genannten subtilen Hinweise aus der Aufgabenstellung in Kombination mit der Erfahrung von Codierern mit eigener Unterrichtspraxis insgesamt zu der aussagekräftigsten Kategorisierung führen. Zudem verspricht die taxonomische Einordnung der Aufgaben auch ohne die Relativierung am vorausgegangenen Unterricht interessante Ergebnisse. Bereits die Anforderungen der Aufgaben, wie sie sich unabhängig von der Bearbeitung und Übung im vorangegangenen Unterricht stellen, vermitteln ein Bild von konkret verlangtem Wissen und Können bzw. von einem Standard, an dem die Maturanden und Maturandinnen gemessen werden. Wie auch von Experten viele Aufgaben, die von Novizen hohe Denkleistungen erfordern, nur noch mittels mechanischem Abrufen von verinnerlichtem Wissen gelöst werden, können gleiche Aufgaben verschiedenen Maturanden je nach bisherigen Erfahrungen im fachlichen und auch im überfachlichen Bereich unterschiedliche Denkleistungen abfordern. Trotzdem ist für alle die konkrete Anforderungssituation dann erfolgreich bewältigt, wenn sie zur gleichen, richtigen Lösung gelangen. Mit der Untersuchung im Teilprojekt D1 werden diese Aufgaben aus einer personenunabhängigen Perspektive charakterisiert. Dabei kann nicht mehr, aber auch nicht weniger als **das Potenzial des kognitiven Anspruchsniveaus** ermittelt werden (siehe ähnlich im COACTIV-Projekt, Jordan et al., 2006). Dasselbe Problem besteht übrigens – wenn auch in abgeschwächter Form – bei zentralen, standardisierten Tests: Je nach Verlauf des Unterrichts können auch Standard-Testfragen für den einen Probanden blosse Reproduktions- und für den anderen schwierige Transferaufgaben bedeuten. Trotzdem ist bereits die schülerunabhängige taxonomische Analyse des Schwierigkeitspotenzials der Aufgaben interessant. Offen bleiben auch Fragen der Vergleichbarkeit des Aufgabenpotenzials mit den tatsächlich in den Prüfungen verlangten Soll-Lösungen und den erreichten Punktzahlen und Noten.

Zusätzlich zur Analyse der kognitiven Anforderungen wurden die Aufgabeninhalte ermittelt und kategorisiert, die im Jahre 2007 Gegenstand von schriftlichen Schweizer Maturitätsprüfungen waren. Ziel war eine an den Kompetenzrastern orientierte, zusammenfassende Beschreibung pro untersuchtes Fach. In den Testfächern Biologie und Mathematik wurden zur Gliederung die bereits erstellten Kompetenzraster verwendet. Sie mussten teilweise ergänzt werden mit der Aufnahme von Inhaltskategorien, die für die Tests von Teilprojekt C nicht vorgesehen waren, aber Gegenstand der untersuchten Maturitätsprüfungen sind.

2 Beschreibung der Stichprobe

Die Stichprobe besteht aus denselben Schulen und Klassen wie in Teilprojekt C (siehe Kapitel IV, Abschnitt 2). Das von den Schulen angeforderte Datenmaterial wurde fast vollständig geliefert. An diesem Teilprojekt haben sich – mit Ausnahme des Liceo cantonale di Bellinzona – auch die Tessiner Gymnasien beteiligt.

3 Ergebnisse

3.1 Erstsprache

Für Erstsprache wurden die konkreten Aufgabenstellungen von insgesamt 122 Schulen (68 in der Deutschschweiz [D], 50 in der Romandie [F] und 4 im Tessin [I]) mit total 314 Klassen (D 106, F 174, I 34) untersucht. Insgesamt lagen für die 314 Klassen 152 unterschiedliche Prüfungen (D 94, F 54, I 4) vor. Es gab also auch klassenübergreifende Prüfungen. Die 314 (teilweise gleichen) Prüfungen, enthalten insgesamt 872 (teilweise gleiche) Aufgabenstellungen (D 446, F 264, I 162).

3.1.1 Das Analyse-Instrumentarium

Um das für Erstsprache geeignetste und breitest akzeptierte Instrumentarium zur Analyse des kognitiven Anforderungspotenzials zu ermitteln, wurden ausgedehnte Literaturrecherchen unternommen und Expertenbefragungen durchgeführt. Als Resultat kann kurz gefasst vermerkt werden, dass sich im Bereich der Erstsprache neben der „alten“ Taxonomie von Bloom et al. (1956) bisher keine neuere und passendere Taxonomie zur Bestimmung der kognitiven Anforderungen von Prüfungsaufgaben durchgesetzt hat. Aus dem englischsprachigen Raum wurde in letzter Zeit eine Adaption der Bloom’schen Taxonomie durch Narciss & Proske (2001) auch in den deutschsprachigen Raum eingeführt. Dieses Modell, das sich vor allem auch auf das universitäre Niveau beziehen soll, wird in laufenden Schweizer Studien bereits eingesetzt und wird sich vermutlich für den Bereich Erstsprache als ausbaufähig erweisen. Diese neuere Narciss’sche Adaption konnte unseres Erachtens das Fehlen einer breiten Akzeptanz nicht durch herausragende Vorzüge gegenüber dem herkömmlichen Instrumentarium wettmachen. Es blieb deshalb auch für das Fach Erstsprache dabei, die revidierte Taxonomie von Anderson et al. (2001) zu verwenden (siehe Abschnitt 1.2, Tab. V.1). Für die Erstsprachen ging es bei der Vorarbeit zur Analyse der kognitiven Anforderungen hauptsächlich darum, die vor allem für Naturwissenschaften formulierten Definitionen und Beispiele der Kategorien und Unterkategorien des Analyserasters auf die Anforderungen und Gegebenheiten von Sprachleistungen zu übertragen.

Zur Analyse der Prüfungsinhalte in Erstsprache wurden folgende Kategorien erfasst:

die Art der verlangten Aufsatzform (Essay, Erörterung, Interpretation o. Ä.)

4. die Textsorte der beigegebenen Vorlage (Gedicht, Erzählung, Sachtext o. Ä.)
5. das Jahrhundert, aus dem die gewählte Textvorlage stammt
6. der Dichter oder die Autorin der Textvorlage, Herkunft und Geschlecht
7. das Thema oder die Themen der Aufgabenstellung bzw. der Textvorlage
8. Des Weiteren wurden auch die folgenden, rein quantitativen Merkmale erfasst:
9. die Länge der vorgelegten Texte
10. die Anzahl der zu bearbeitenden Vorlagen/Primärtexte und der vorgegebenen Auswahlmöglichkeiten

Als weitere Merkmale der Prüfungen wurden erfasst:

die Art und Anzahl der unterstützenden Hinweise (Worterklärungen, Kontext o. Ä.)

11. die Art und Anzahl möglicher Hilfsmittel (Duden, PC o. Ä.)
12. die Dauer der Prüfung (wo angegeben)

Im Folgenden werden die Ergebnisse der Analyse in der Reihenfolge der einzelnen Punkte dargestellt.

3.1.2 Die Prüfungsinhalte in Erstsprache

3.1.2.1 Die Aufgabenform

Grundsätzlich ist die Art der gestellten Aufgabe in jeder – untersuchten – schriftlichen Erstsprache-Maturaprüfung unseres Landes einheitlich: Es handelt sich immer um den Auftrag, einen schriftlichen Aufsatz von mehreren Seiten zu produzieren. In keinem der Fälle werden reine Grammatikaufgaben oder direkte Verständnis- und Wissensfragen zu einem Teilgebiet der Literatur, der Literaturtheorie und -geschichte, der Kultur- oder Sprachgeschichte gestellt, weder in Deutsch noch in Französisch oder Italienisch als Erstsprache. In jedem Fall gilt es für die Schweizer Maturandinnen und Maturanden, ein selbstständig durchdachtes Produkt zu kreieren, in der Regel im Rahmen von vier Stunden, zumeist als Reinschrift von Hand oder selten unter Zuhilfenahme eines eigenen oder schulinternen Laptops/Computers. Unter dieser oberflächlichen Einheitlichkeit verbirgt sich bei genauerem Hinsehen jedoch ein breites und interessantes Feld von Vielfalt und Unterschiedlichkeiten. Sie sind für die Deutschschweiz aus Tabelle V.2 ersichtlich.

Tab. V.2: Die Matura-Aufsatzformen an den Deutschschweizer Gymnasien

Aufsatzformen Deutsch	Erstwahl	Zweitwahl	Drittwahl	Viertwahl	Total (N)	Gruppen (N)	Häufigkeit (%)	Gruppen (%)
Interpretation	54	2	0	0	56		10.6%	
Vergleichende Interpretation	14	0	0	0	14		2.6%	
Analyse	13	0	0	0	13	83	2.5%	16%
Erörterung (freie)	73	6	0	0	79		15.0%	
Erörterung (literarische)	18	4	0	0	22		4.2%	
Erörterung (textgebundene)	19	0	0	0	19		3.6%	
Argumentation	6	2	0	0	8	128	1.5%	24%
Essay	25	16	3	0	44		8.3%	
Betrachtung	5	0	0	0	5	49	0.9%	9%
Kurzgeschichte, Erzählung	21	8	0	1	30		5.7%	
Glosse	0	2	0	0	2		0.4%	
Satire	2	0	1	0	3		0.6%	
Groteske	1	0	0	0	1		0.2%	
Märchen/Parabel	0	4	0	0	4		0.8%	
Gedicht	0	1	0	2	3		0.6%	
Szene	0	0	3	0	3	46	0.6%	9%
Rede	5	1	1	1	8		1.5%	
Brief	0	7	0	0	7		1.3%	
Zeitungsartikel	2	1	0	0	3		0.6%	
Dialog	0	0	4	1	5		0.9%	
Tagebuch	0	0	0	4	4	27	0.7%	5%
freie Aufsatzform	188	4	3	0	195	195	36.9%	37%
Total Aufgabenarten	446	58	15	9	528		100.0%	100%
keine Wahl	133				133		29.8%	

Insgesamt machen weit mehr als die Hälfte der Lehrpersonen (D 63%) klare Formvorgaben für das Endprodukt Aufsatz: In rund einem Viertel aller Aufsätze wird eine argumentative Textform vorgeschrieben (Erörterung, Argumentation), für einen Sechstel der Aufsätze wird eine Interpretation oder Analyse eines Textes verlangt, immerhin in knapp einem Zehntel aller Aufsätze wird die hochstehende Betrachtungsweise eines Essays erwartet, und nur bei knapp einem Zehntel aller Aufsatzthemen sind auch künstlerisch anspruchsvolle Erzählformen erlaubt. Erstaunlich wenige Lehrpersonen geben ihren Maturandinnen und Maturanden auch die Gelegenheit, einen fiktiven Zeitungsartikel, eine Rede oder einen Brief zu formulieren. Obwohl bei den meisten Aufgabenstellungen eine Zweit- oder gar Dritt- und Viertwahl bezüglich der Aufsatzform möglich ist, wird doch für rund einen Viertel aller Aufsätze die Form strikt vorgegeben. 37% der Deutschschweizer Maturandinnen und Maturanden können eigenständig und ohne Einschränkungen und Wahlvorschläge selbst die Form ihres Aufsatzes bestimmen (so genannte „freie Aufsatzform“), das heisst in diesen Fällen macht die Lehrperson keine expliziten Formvorgaben, so dass die angemessene Form selber aus der Art der Aufgabenstellung oder der Textvorlage erschlossen werden muss.

Die in der italienischsprachigen Schweiz gefundenen Aufsatzformen sind in Tabelle V.3 dargestellt. Hier wird noch häufiger eine freie Aufsatzform (componimento) zur Verfügung gestellt (48%). In 52% der Prüfungen machen die Lehrpersonen eine klare Formvorgabe: In den meisten Fällen wird hier eine Analyse erwartet (40%), in 12% sind auch andere Formen verlangt (riflessione, confronto di testi). Bei keiner der italienischsprachigen Aufgabenstellungen mit Formvorgaben gibt es eine Wahlmöglichkeit zwischen verschiedenen Aufsatzformen.

Tab. V.3: Die Matura-Aufsatzformen in der italienischsprachigen Schweiz

Aufsatzformen Italienisch	Erstwahl	Häufigkeit (%)
Freie Aufgabensorte (componimento)	78	48%
Analisi (poetico)	34	21%
Analisi (narrativo)	31	19%
Confronto di testi (Vergleichende Interpretation)	11	7%
Riflessione (Betrachtung)	8	5%
Total Aufgabenarten	**162**	100%
keine (Zweit-)Wahl	162	100%

In der französischsprachigen Schweiz (siehe Tabelle V.4) ergeben sich bezüglich der verlangten Aufsatzformen folgende Resultate: In rund der Hälfte der Fälle ist die geforderte Form eine „dissertation“, die sich entweder auf literarische oder eher generelle Themen bezieht und in etwa der deutschsprachigen „Erörterung“ entspricht. Insgesamt gibt es zwölf verschiedene Aufsatzformen. Neben dem „commentaire composé (littéraire)“ (rund 9%), die der deutschsprachigen „Interpretation“ entspricht, und dem argumentativen Text (rund 2%) ragt nur die Sorte „evocation“ heraus (5%). Es handelt sich um eine sogenannte „Erweckung“, in der die von poetischen Texten ausgelösten Gefühle, Stimmungen und Bilder in künstlerisch-bildhafter Sprache dargestellt werden sollen. Dies ist eine Aufsatzprüfungsform, die in den anderen beiden Sprachregionen so nicht vorkommt und eine Besonderheit der französischen Sprachregion darstellt.

Tab. V.4: Die Matura-Aufsatzformen in der französischsprachigen Schweiz

Type d'exercice	Entsprechende deutschsprachige Aufsatzformen	Erstwahl	Zweitwahl	Häufigkeit (%)
Etude de texte	Analyse	1		0.4%
Commentaire composé (littéraire)	Interpretation (immer literarisch)	24		9.1%
Comparaison de textes	Vergleichende Interpretation	6		2.3%
Dissertation Sujet littéraire	Erörterung literarische	71		26.9%
Dissertation libre	Erörterung (freie)	0		0.0%
Dissertation Sujet général	Erörterung (textgebundene)	62		23.5%
Texte argumentatif	Argumentation	6		2.3%
Essai	Essay	0		0.0%
Narration	Kurzgeschichte, Erzählung	1		0.4%
Commentaire de texte		5	1	1.89%
Evocation		13	15	4.9%
Méditation libre	freie Aufsatzform	0	1	0.0%
Texte libre	freie Aufsatzform	1	1	0.4%
Sans indication	freie Aufsatzform	74		28.0%
Total Aufgabenarten		264	18	100.0%
keine Wahl			246	93.2%

Kurz gefasst kann gesagt werden, dass sich die Situation bezüglich der verlangten Aufsatzformen in den drei Sprachregionen sehr unterschiedlich darstellt: Während in der italienischsprachigen Schweiz fünf Aufsatzformen geprüft werden und häufig eine freie Form möglich ist, werden in der französischsprachigen Schweiz zwölf und in der deutschsprachigen Schweiz 22 Aufsatzformen geprüft. In der Romandie wird am häufigsten eine klare Aufsatzform vorgegeben (72%), und dort ist ähnlich wie im Tessin (I 100%) oft keine Alternative im Sinn einer Zweitwahl möglich (F 93%). Im Unterschied dazu können (oder müssen) in der Deutschschweiz häufig (bei 313 von 446 Aufgabenstellungen, also 70%) Prüflinge eine Entscheidung zwischen verschiedenen möglichen Aufsatzformen treffen oder erhalten diesbezüglich keine Vorgabe.

3.1.2.2 Der vorgelegte Primärtext

Bei einem Primärtext handelt es sich um kürzere Zitate oder längere Textauszüge, welche der Formulierung der eigentlichen Aufgabenstellung vorangehen, ihr folgen oder allenfalls als Zusatzseiten beigeheftet sind. Die Maturandinnen und Maturanden haben die Aufgabe, sich in ihrem Produkt mit dieser Vorlage auseinanderzusetzen und sich in ihren Aufsätzen darauf zu beziehen.

Von den zitierten Primärtexten stammen die meisten aus dem Bereich „Literarische Texte" (D 41%, F 62%, I 57%). Sachtexte bilden den weitaus kleineren Teil (D 20%, F 21.7%, I 9%) und weitere Vorlagen sind der Kategorie „Philosophische Texte" (D 12%, F 14%, I 28%) zuzuordnen. Wenige der Vorlagen sind Bilder, Zeichnungen oder Karikaturen (D 4%, F 1%; I 0%), wobei im Tessin auch ergänzende Illustrationen (z.B. Autorenporträt) verwendet werden (I 14%). Als regionale Besonderheit wird in der Deutschschweiz ein grosser Teil der Vorlagen,

insgesamt 16%, als Text vorgelegt, welchen die Lehrpersonen selber verfasst haben, so etwa überblicksartige Zusammenfassungen mit Kommentaren zum Zeitgeschehen oder moralisch-ethische Fragen und Gedanken zu gesellschaftlichen Phänomenen (F 2%, I 6%). (Bei einem Rest von 6% in der Deutschschweiz konnte die Textsorte nicht eindeutig zugeordnet werden.)

Eine detaillierte Gliederung für die Deutschschweiz, die Romandie und das Tessin findet sich in den Tabellen V.5, V.6 und V.7.

Tab. V.5: Die Textsorten der Vorlagen der Deutschschweizer Maturitätsprüfungen

Primärtexte Deutsch: Textsorten		Anzahl (N)	Häufigkeit (%)	Gruppe (%)
Literarische Texte	Aphorismus	53	26%	
	Anekdote, Parabel	4	2%	
	Gedicht	70	34%	
	Erzählung, Novelle	40	20%	
	Roman	8	4%	
	Drama	13	6%	
	Briefwechsel	8	4%	
	Autobiografie/Interview	7	3%	
Total Literarische Texte		203	100%	41%
Philosophische Texte	Aphorismus	45	76%	
	Abhandlung	14	24%	
Total Philosophische Texte		59	100%	12%
Sachtexte	Leserbrief	3	3%	
	Referat/Rede	7	7%	
	Essay	25	25%	
	Zeitungsartikel	37	37%	
	Reklame	6	6%	
	Sachbuch	14	14%	
	wissenschaftliches Fachbuch	4	4%	
	Website	4	4%	
Total Sachtexte		100	100%	20%
Lehrertexte	Fragen	22	28%	
	Kommentar	57	72%	
Total Lehrertexte		79	100%	16%
Bilder	Fotografie	6		
	Gemälde	2		
	Zeichnung/Grafik	1		
	Karikatur/Cartoon/Comic	10		
Total Bilder	Bild ohne Text/Titel	21		4%
Textsorte nicht eindeutig		30		6%
Total Texte (alle Textsorten)		492		100%

Den Tabellen kann entnommen werden, wie sich innerhalb der beschriebenen Bereiche die einzelnen Textsorten verteilen: So sind beispielsweise in der Deutschschweiz mehr als ein Drittel der literarischen Primärtexte Gedichte, während nur 6% der Literaturvorlagen aus Theaterstücken entnommen wurden; bei den Sachtexten sind 37% Zeitungsartikel und nur 4% der Sachtexte wurden offenkundig einer Website entnommen.

In der Romandie sind mehr als ein Drittel der literarischen Vorlagen Auszüge aus Romanen, je ein knappes Viertel sind Gedichte und Essays. Zeitungsartikel werden nur zu 13% zitiert.

Tab. V.6: Die Textsorten der Maturitätsprüfungen aus der Romandie

Primärtexte Französisch: Textsorten		Anzahl (N)	Häufigkeit (%)	Gruppe (%)
Textes littéraires				
	roman	18	9.7%	
	aphorisme	42	22.6%	
	poème	11	5.9%	
	récit	67	36.0%	
	drame	5	2.7%	
	essai	42	22.6%	
	correspondance	1	0.5%	
	autobiographie	0	0.0%	
Total „Textes littéraires“		186	100.0%	62.0%
Textes philosophiques				
	aphorisme	20	46.5%	
	traité	23	53.5%	
Total „Textes philosophiques“		43	100.0%	14.3%
Autres textes				
	lettre de lecteur/lectrice	0	0.0%	
	exposé/discours	0	0.0%	
	essai	56	86.2%	
	ouvrage spécialisé	0	0.0%	
	ouvrage de vulgarisation	0	0.0%	
	article de journal	9	13.8%	
	publicité	0	0.0%	
Total „Autres textes“		65	100.0%	21.7%
question enseigant-e		6		
commentaire enseignant-e		0		
Total „Textes enseignant-e“		6		2.0%
Total textes (tous les genres)		300		100.0%

Im Tessin liegt das Schwergewicht auf Gedichten (20%) und Erzählungen (13%) bzw. auf den philosophischen Vorlagen (28%).

Tab. V.7: Die Textsorten der Maturitätsprüfungen aus der italienischsprachigen Schweiz (Tessin)

Primärtexte Italienisch: Textsorten		Anzahl (N)	Häufigkeit (%)	Gruppe (%)
Literarische Texte				
	Aphorismus	0	0%	
	Gedicht	34	20%	
	Erzählung	23	13%	
	Roman (2 x Manzoni)	11	6%	
	Novelle	8	5%	
	Drama	0	0%	
	Briefwechsel	0	0%	
	Autobiografie	11	6%	
	Artikel über einen Roman	11	6%	
Total Literarische Texte		98	57%	57%
Philosophische Texte				
	Aphorismus	7	4%	
	Abhandlung (discussione), Argumentation	15	9%	
	Betrachtung	16	9%	
	Rede	11	6%	
Total Philosophische Texte		49	28%	28%
Sachtexte				
	Leserbrief	0	0%	
	Referat/Rede	0	0%	
	Essay	8	5%	
	Fachbuch	0	0%	
	wissenschaftliches Fachbuch (historischer Text)	7	4%	
	Zeitungsartikel	0	0%	
	Reklame	0	0%	
Total Sachtexte		15	9%	9%
Lehrertexte				
	Fragen	0	0%	
	Kommentar	11	6%	
Total Lehrertexte		**11**	**6%**	6%
Textsorte nicht eindeutig		0	0%	
Total Texte (alle Textsorten)		173		100%

3.1.2.3 Zeitbezug

Bei allen untersuchten Schweizer Prüfungen (siehe Tab. V.8) geht ein starker Trend in Richtung neuere Texte; aus dem 20. Jahrhundert stammen insgesamt die meisten Vorlagen (D 37%, F 65%, I 39%), gefolgt von noch aktuelleren Texten aus dem 21. Jahrhundert (D 42%, F 14%, I 9%). Nur noch selten werden Maturandinnen und

Maturanden in der Maturitätsprüfung mit Auszügen aus der Literatur des 17. bis 19. Jahrhunderts konfrontiert (D 14%, F 15%, I 27%). Die Ausnahme bildet das Tessin, bei dem die Vorlagen aus der Zeit zwischen dem 14. und 16. Jahrhundert 20% aller Texte ausmachen, deutlich mehr als die Texte aus dem 21. Jahrhundert (9%).

Tab. V.8: Angaben, aus welchem Jahrhundert die Vorlagen stammen (Schweizer Maturitätsprüfungen, alle drei Sprachregionen)

Jahrhundert	D (N)	D (%)	D (%) Gesamt	F (N)	F (%)	F (%) Gesamt	I (N)	I (%)	I (%) Gesamt
Antike/vor Christus	7	1,5%		0	0.0%		8	4.6%	
1.–13. Jh.	2	0.4%		0	0.0%		0	0.0%	
14.–16. Jh.	2	0.4%	0.8%	9	3.0%	3.0%	34	19.7%	24.3%
17. Jh.	8	1.7%		6	2.0%		14	8.1%	
18. Jh.	26	5.4%		4	1.3%		11	6.4%	
19. Jh.	34	7.1%	14.1%	36	11.8%	15.1%	22	12.7%	27.2%
frühes 20. Jh.	83	17.3%		73	23.9%		23	13.3%	
spätes 20. Jh.	96	20.0%	37.2%	126	41.3%	65.2%	45	26.0%	39.3%
21. Jh.	121	25.2%		44	14.4%		16	9.2%	
21. Jh. (Lehrperson)	81	16.8%	42.0%			14.4%	0	0.0%	9.2%
keine Angabe	21	4.4%		7	2.3%		0	0.0%	
Total erfasst	481	100.0%		305	100.0%		173	100.0%	

3.1.2.4 Autorinnen und Autoren

Dem festgestellten Zeittrend zur Moderne und Postmoderne entspricht auch die lange Liste der von den Lehrpersonen in ihren Prüfungsthemen zitierten Dichterinnen, Schriftsteller, Dramatiker und Essayisten: Vom früheren klassischen Kanon des Literaturunterrichts ragen in der Deutschschweiz nur Goethe, Brecht, Kafka, Fried, Rilke und Dürrenmatt mit je mehr als fünf Zitaten aus dem breiten Feld der Autoren heraus; sie stellen knapp ein Viertel aller literarischen Vorlagen (siehe im Einzelnen in Tabelle V.9).

Tab. V.9: Zitierte Autorinnen und Autoren (Auszug) in Deutschschweizer Maturitätsprüfungen

Autor/Autorin Primärtext Deutsch	Anzahl (N)	Insgesamt (%)	Literarisch (%)
Lehrperson	77	15.7%	
Journalist/in	36	7.3%	
unbekannt (allgemeine Lebensweisheiten u. Ä.)	15	3.0%	
kein Primärtext	45	9.1%	
keine Angabe	19	3.9%	
Ausländer, Rose	2	0.4%	1.0%
Bachmann, Ingeborg	2	0.4%	1.0%
Benn, Gottfried	5	1.0%	2.5%
Böll, Heinrich	1	0.2%	0.5%
Brecht	11	2.2%	5.4%
Domin, Hilde	2	0.4%	1.0%
Dürrenmatt	7	1.4%	3.4%
Eichendorff	2	0.4%	1.0%
Fried, Erich	7	1.4%	3.4%
Goethe	11	2.2%	5.4%
Hahn, Ulla	5	1.0%	2.5%
Handke	4	0.8%	2.0%
Hofmannsthal	2	0.4%	1.0%
Kästner, Erich	3	0.6%	1.5%
Kafka	8	1.6%	3.9%
Kaschnitz, Marie Luise	3	0.6%	1.5%
Kleist	2	0.4%	1.0%
Kunert, Günther	4	0.8%	2.0%
Lessing	3	0.6%	1.5%
Mann, Thomas	2	0.4%	1.0%
Nietzsche	3	0.6%	1.5%
Novalis	3	0.6%	1.5%
Rilke	6	1.2%	3.0%
Schiller	3	0.6%	1.5%
Tucholsky	3	0.6%	1.5%
Biermann	2	0.4%	1.0%
Grönemeyer	1	0.2%	0.5%
Bundesrat Blocher	3	0.6%	
Bundesrat Couchepin	1	0.2%	
Bundesrat Leuenberger	1	0.2%	
Total (Auszug)	304	59.6%	52.7%
Total Autor/innen Deutsch	492	100.0%	

In einer Hinsicht bleiben die Deutschschweizer Lehrpersonen dem traditionellen Gedanken der Germanistik treu: Sie zitieren in ihren Vorlagen weitgehend original deutschsprachige Literatur bzw. Autorinnen und Autoren des deutschen Sprachraums (siehe Tabelle V.10). Von den rund 82% Primärtexten aus „deutscher Schreibhand" stammen 14% aus der einheimischen Schweizer Literatur und Presse. Übersetzungen haben auch in unserem globalisierten Zeitalter weiterhin nur einen geringen Stellenwert in Deutschprüfungen: Ganz wenige Texte stammen aus dem frankophonen, asiatischen oder anglophonen Raum.

Tab. V.10: Herkunft/Nationalität der Autorinnen und Autoren (Deutschschweizer Maturitätsprüfungen)

Herkunft	Anzahl (N)	Häufigkeit (%)	Gruppe (%)
Schweiz	69	14.0%	
vermutlich Schweiz (Lehrperson/Journalist/in)	100	20.3%	
Deutschland	185	37.6%	
Österreich	36	7.3%	
Tschechoslowakei (deutschsprachig)	13	2.6%	81.9%
Frankreich	13	2.6%	
Italien	7	1.4%	
englischsprachiger Raum	21	4.3%	
spanische Literatur	3	0.6%	
russische Literatur	4	0.8%	
Asien	9	1.8%	
weitere	7	1.4%	13.0%
unbekannt	25	5.1%	5.1%
Total Autor/innen Deutsch	492	100.0%	100.0%

Ein ähnlicher Bezug zeigt sich in der Romandie (siehe Tabelle V.11): Die direkten Zitate aus der französischsprachigen Literatur stellen den grössten Anteil (84%), Übersetzungen aus anderen Sprachen sind selten. Einheimische Schweizer Literatur macht in der Romandie rund 10% aus, davon sind vermutlich 3 von 30 Texten Übersetzungen.

Tab. V.11: Herkunft/Nationalität der Autorinnen und Autoren (Französischsprachige Maturitätsprüfungen aus der Romandie)

Nationalité de l'auteur-e	Anzahl (N)	Häufigkeit (%)	Gruppe (%)
CH	24	7.9%	
CH probablement (ens./journaliste)	6	2.0%	9.9%
D	5	1.7%	
A	7	2.3%	4.0%
F/pays de l'est	36	11.9%	
F	188	62.3%	84.1%
I	5	1.7%	
anglophone	2	0.7%	
hispanophone	0	0.0%	
russophone	0	0.0%	
Asie	0	0.0%	
autre	4	1.3%	3.6%
inconnu	25	8.3%	8.3%
total	302	100.0%	100.0%

In der italienischen Schweiz stammen 17 der 20[30] zitierten Autorinnen und Autoren aus Italien; ein Autor stammt aus dem frankophonen, einer aus dem anglophonen Raum und einer aus dem frühen Griechenland (Aristoteles). Im Tessin wurde keine Vorlage aus der Schweizer Literatur verwendet.

Für alle drei Sprachregionen der Schweiz kann somit festgestellt werden, dass hauptsächlich Literatur in Originalsprache verwendet wird (D 82%, F 84%, I 85%), Übersetzungen aus benachbarten Sprachen bleiben die Ausnahme. Speziell auf die Schweizer Literatur beziehen sich hauptsächlich die Deutschschweizer Lehrpersonen (14%), gefolgt von jenen in der Romandie (9.9%).

Auf einer weiteren sehr traditionellen Schiene bleiben die Schweizer Lehrpersonen in ihrer eindeutig nicht gendergerechten Auswahl der Primärtexte (siehe Tabelle V.12): In allen drei Sprachregionen schaffen es unverhältnismässig wenige Texte mit weiblicher Autorschaft, von den Lehrpersonen zitiert zu werden. In der Deutschschweiz bleiben es lediglich 9% Autorinnen, in der Romandie 10%, im Tessin 5%. Hier zeigt sich im Sinn eines zeitgemässen gendergerechten Unterrichts ein deutlicher Aufhol- und Aufklärungsbedarf der Lehrerschaft aller drei Sprachregionen.

Tab. V.12: Gender/Geschlecht der Autorinnen und Autoren (Vorlagen an Schweizer Maturitätsprüfungen)

Gender der Autor/innen	D (N)	D (%)	F (N)	F (%)	I (N)	I (%)
weiblich	40	8.7%	31	10.2%	8	5%
männlich	322	69.7%	266	87.2%	165	95%
unbekannt (Lehrperson/Journalist/in)	100	21.6%	8	2.6%	0	0%
Total erfasst	462	100.0%	305	100.0%	173	100%

[30] Die Zählung fürs Tessin basiert hier auf der Ebene aller Prüfungen (I 4), welche in allen vier Schulen klassenübergreifend eingesetzt wurden.

3.1.2.5 Themen

Von den 446 möglichen Aufsatzthemen der untersuchten Deutschschweizer Maturitätsprüfungen konnten rund 300 Themen als Stichworte erfasst werden. Sie wurden den drei grossen Textsorten entsprechend gruppiert in literarische Themen, philosophische Themen und Sachthemen. Bei den vorgelegten Themen können die folgenden Auffälligkeiten hervorgehoben werden: Im Bereich der literarischen Themen finden sich klassische Literaturthemen wie Liebe, Sehnsucht, Krise, Krieg, Natur, Stadt, Schicksal, Zufall, Charakter, Reisen und viele weitere. Daneben zeigt sich eine auffallende Häufung in den Meta-Bereichen „Funktion von Literatur/Wirkung von Büchern/Lesen als Abenteuer" sowie „Sprache und Wirklichkeit". Eng literaturwissenschaftliche oder -geschichtliche Fragestellungen – beispielsweise zur Gattungstheorie, zu politischer Lyrik, epischem Theater oder Verfilmungen – sind auffallend selten. Bei den philosophischen Themen tauchen – auszugsweise – Fragen auf zur Verantwortlichkeit des Menschen, seiner (selbstverschuldeten) Unmündigkeit, zu Mut, Glück und Mitleiden, zu Heimat und dem Fremden, zu Vergangenheit und Zukunft, zur Funktion von Bildung, Ordnung und der „Weisheit" der Mehrheit. Die Sachthemen sind meist aktuell und breit gestreut: von Atomkraftwerken und dem Klimabericht über die Rolle des Staates bis hin zu Religion, Erziehung, Mindestlohn, Spitzenleistungen, Sterbehilfe, Demokratie und vielem mehr. Ein eigenes grosses Feld machen Themen aus der Lebenswelt der Jugendlichen aus: die Macht der Computer, virtuelle Welt, Handy, Rausch und Alkohol, Jugendkriminalität, Anglizismen in der Sprache, Traumberuf Model, Verschuldung, Stimm- und Wahlrecht ab 16 Jahren und ähnliche.

Insgesamt macht das Feld der an den Deutschschweizer Maturitätsprüfungen zu bearbeitenden Aufsatzthemen inhaltlich einen ansprechenden, vielfältigen und aktuellen Eindruck. Die Maturandinnen und Maturanden setzen sich in ihren Abschlussaufsätzen mit lebensphilosophischen, ethisch relevanten Fragen ihrer gegenwärtigen und zukünftigen Lebenswelt auseinander, machen sich Gedanken zu Gegenwartsphänomenen sowie zu zeitlosen Aspekten des Menschseins.

3.1.2.6 Länge der vorgelegten Texte

Als weiteres Merkmal, das nebst dem Inhalt auch den Anforderungsgrad einer Prüfung bestimmt, kann die Länge bzw. die Quantität der vorgelegten Texte ermittelt werden: So macht es einen Unterschied, ob seitenlange dialogische Auszüge aus einem Theaterstück rezipiert werden müssen, ob ein kurz gefasster Aphorismus in zwei Zeilen verstanden werden will oder ob ein 40-Zeilen-Gedicht mit dem Inhalt einer 100-zeiligen Kurzgeschichte verglichen werden muss.

Auch bezüglich der Textlänge ergibt sich ein erstaunlich heterogenes Ergebnis, sowohl zwischen den Sprachregionen als auch innerhalb der Sprachgemeinschaften (siehe Tabelle V.13). So werden in der Deutschschweiz immerhin in rund einem Fünftel aller Prüfungen die Maturandinnen und Maturanden von ihren Lehrpersonen mit längeren, bis 100-zeiligen, oder sehr langen, über 100-zeiligen Texten konfrontiert. Auf der anderen Seite legt man etwa der Hälfte aller Deutschschweizer Gymnasiastinnen und Gymnasiasten nur ein- bis fünfzeilige Kurztexte vor. Nur gut ein Viertel aller deutschsprachigen Prüflinge beschäftigt sich mit mittleren, 6- bis 20-zeiligen Texten. In der Romandie werden zwei Dritteln der Maturandinnen und Maturanden kurze, ein- bis fünfzeilige Texte vorgelegt, ein Achtel der Prüflinge beschäftigt sich mit Texten mittlerer Länge, und nur rund 8% müssen auch an langen bis sehr langen Texten arbeiten.

Tab. V.13: Länge der Vorlagentexte (Schriftliche Maturitätsprüfungen, alle drei Sprachregionen)

Primärtext	Länge	D (N)	D (%)	D Gruppe (%)	F (N)	F (%)	F Gruppe (%)	I (N)	I (%)	I Gruppe (%)
ohne	nur Stichwort/nur Titel	45	9%		0	0%		0	0%	
kurz	1–2 Zeilen	129	26%		144	47.2%		18	10%	
	3–5 Zeilen	96	20%	46%	59	19.3%	66.5%	11	6%	17%
mittel	6–10 Zeilen	73	15%		13	4.3%		18	10%	
	11–20 Zeilen	66	13%	28%	25	8.2%	12.5%	50	29%	39%
lang	21–40 Zeilen	48	10%		8	2.6%		23	13%	
	mehr als 40 Zeilen	42	8.5%	18.5%	13	4.3%	6.9%	30	17%	30.5%
sehr lang	mehr als 100 Zeilen	10	2%		5	1.7%		23	13%	
	ganzes Werk	7	1.5%	3.5%	34	11.2%	12.8%	0	0%	13.5%
Bild	nur Bild, ohne Textvorlage	21	4%	4%	3	1%	1%	0	0%	
	texte manque	0	0%		1	0.3%		0	0%	
	Total Texte (Länge)	492	100%		305	100%		173	100%	

Viel einheitlicher verhalten sich die Schweizer Lehrpersonen bezüglich der Anzahl der vorgelegten Primärtexte (siehe Tabelle V.14): In der Mehrheit aller Prüfungen liegt ausschliesslich ein Primärtext vor (D 75%, F 87%, I 93%), zwei Vorlagetexte finden sich bei etwa einem Achtel der Prüfungsthemen (D 12%, F 13%, I 7%); nur 3% aller Deutschschweizer Prüfungen beruhen auf drei oder mehr Texten, die miteinander verglichen werden müssen.

Tab. V.14: Quantität der Vorlagentexte (Schweizer Maturitätsprüfungen, alle drei Sprachregionen)

Anzahl Texte/Bilder	D (N)	D (%)	F (N)	F (%)	I (N)	I (%)
Keine Primärtexte/Bilder gegeben	45	10%	0	0%	0	0%
Ein Primärtext/Bild gegeben	333	75%	233	86.6%	151	93%
Zwei Primärtexte/Bilder gegeben	53	12%	36	13.4%	11	7%
Drei Primärtexte/Bilder gegeben	11	2%	0	0%	0	0%
Vier Primärtexte/Bilder gegeben	0	0%	0	0%	0	0%
Fünf Primärtexte/Bilder gegeben	4	1%	0	0%	0	0%
Total (Aufgabenstellungen)	446	100%	269	100%	162	100%
Total Anzahl (Primärtexte)	492		305		173	

Als einzige Auffälligkeit beinhalten 10% der Deutschschweizer Prüfungen keine Textvorlage. Hier besteht die Aufgabenstellung darin, auf ein Stichwort hin einen Aufsatz zu schreiben (zum Beispiel zu „Schein“, „Abseits“, „insgeheim“, „Grenzen“, „Schwelle“).

Es kann also bei Schweizer Maturaprüfungen vorkommen, dass der Prüfling ein DIN A4-Blatt mit drei Themen inklusive den zugehörigen Kurztexten erhält und innert kürzester Zeit eine Auswahl treffen sowie sich auf die eigentliche Produktionsaufgabe konzentrieren kann. Es ist aber auch möglich, dass er ein Paket von zehn Seiten vor sich liegen hat, das eine Mischung von äusserst kurzen Zitaten und seitenlangen Beilagen sowie sieben verschiedenen Varianten von Themen enthält, so dass seine selektive Lesekompetenz ganz zu Beginn der Prüfung hochgradig gefordert ist, da er selber den Aufgabenrahmen stecken und eine Wahl treffen muss (siehe auch nächster Abschnitt).

3.1.2.7 Wahlmöglichkeiten

Von 152 untersuchten Schweizer Prüfungen boten lediglich vier (in vier unterschiedlichen Schulen und Kantonen) bloss zwei Themen zur Auswahl (siehe Tabelle V.15); eine Klasse erhielt sogar neun Aufsatzthemen zur Auswahl, siebenmal wurden acht Themen zur Wahl gegeben (in der Romandie). Dies sind die Pole der Wahlmöglichkeiten. Eine gute Mehrheit von rund 71% aller Klassen erhielt die Wahlmöglichkeit zwischen vier oder fünf Themen, im Tessin sogar alle untersuchten Klassen.

Tab. V.15: Auswahlmöglichkeiten an Aufsatzthemen (Maturitätsprüfungen Deutschschweiz und Romandie)

Anzahl Themen zur Wahl	Prüfungen D (N)	Themen D (N)	D (%)	Prüfungen F (N)	Themen F (N)	F (%)
Zwei	3	6	3%	1	2	2%
Drei	19	57	20%	13	39	24%
Vier	39	156	41%	16	64	30%
Fünf	26	130	28%	4	20	7%
Sechs	3	18	3%	8	48	15%
Sieben	3	21	3%	5	35	9%
Acht	0	0	0%	7	56	13%
Neun	1	9	1%	0	0	0%
Total	94	397	100%	54	264	100%

3.1.2.8 Unterstützende Hinweise

Eine ziemlich hohe Übereinstimmung findet sich in der Handhabung von erklärenden oder unterstützenden Hinweisen durch die Lehrpersonen in der Deutschschweiz und der Romandie. In den meisten Fällen werden keine verdeutlichenden Hinweise formuliert (D 73.2%, F 74%). Im Unterschied dazu werden im Tessin relativ häufig Hinweise gegeben (47%), oft in Form von Fussnoten (Worterklärungen, zur Epoche, zum Kontext).

In der Deutschschweiz (siehe Tabelle V.16) finden sich nur in wenigen Prüfungen kurze Hinweise mit Worterklärungen (5.2%, z. B. bei „Editorial“, „Impetus“, „Konsens“, „Nostalgie“), Kommentare zum Kontext (7,3%, z. B. „Reaktion auf einen Leserbrief vom Vortag“), zur vorgelegten Textsorte oder zum Inhalt (z. B. „Bundesrat-Rede zum sozialen Handeln“). Selten und zu vernachlässigen sind lange Inhaltshinweise (0.4%), Hinweise zum verlangten Aufgabentyp (z. B. „Kriterien für kreative Texte sind Originalität, Fantasie“) oder zur Epochenbestimmung (1.3%)[31].

Auch erläuternde Kommentare zum Autor oder zur Autorin finden sich nicht sehr häufig (D 31.5%, I 10%; F nicht erfasst); diese Zusätze beschränken sich in der Deutschschweiz meist auf die Lebensdaten, Berufs- und Staats-

[31] Für die Romandie und das Tessin wurden die Hinweise weniger differenziert erfasst.

bezeichnung (z. B. „französischer Staatskritiker, österreichische Lyrikerin, deutscher Philosoph“), und nur in vier Fällen (D 0.8%) findet sich eine kurze Autobiografie oder eine Abbildung des Autors. Demgegenüber wird in der italienischen Schweiz dem Primärtext eher eine ausführliche Biografie des Autors mit Bild, Daten, Entwicklung und Positionierung seiner Werke innerhalb der Epoche beigefügt (10%).

Erstaunlicherweise war nur in acht Prüfungen die erwartete Aufsatzlänge vorgegeben.

Tab. V.16: Erläuternde Hinweise der Lehrperson (Maturitätsprüfungen der Deutschschweiz)

Hinweise	D (N)	D (%)
Worterklärungen	28	5.2%
Kontext	39	7.3%
Epoche	7	1.3%
Inhaltlich, kurz	32	6.0%
inhaltlich, lang	2	0.4%
zum Aufgabentyp	9	1.7%
zur Textsorte	27	5.0%
keine	393	73.2%
Total Hinweise	537	100.0%

3.1.2.9 Hilfsmittel

Die Benutzung von erlaubten Hilfsmitteln ist in vielen der untersuchten Prüfungen nicht thematisiert (D 42%, F 74%). Vermutlich wurden entsprechende Vorgaben bereits vor der Prüfung gemacht. Bei den Deutschschweizer Prüfungen, die explizite Angaben enthalten, ist es etwa in der Hälfte der Klassen (53%) erlaubt, den einbändigen „Duden für die deutsche Rechtschreibung“ zu benutzen (allenfalls auch ein anderes Wörterbuch wie „Wahrig“ oder „Bertelsmann“). Sehr selten sind auch zwei Wörterbücher erlaubt oder etwa der Duden und das Computer-Rechtschreibprogramm (4%). Nur äusserst selten werden explizit jegliche Hilfsmittel verboten (2%). Nur in einer der Prüfungen wurde vorgeschrieben, ein bestimmtes Werk der Schullektüre mitzubringen. Interessanterweise gestaltet sich in einigen Schulen mit mehreren Klassen die Regelung der Hilfsmittel nicht einheitlich.

Im Unterschied zur Deutschschweiz kann bzw. muss in der Romandie bei einem Viertel der Prüfungen die Schullektüre (oeuvres) mitgebracht werden (24%). Hier erwartet die Lehrperson offensichtlich, dass die Maturandinnen und Maturanden auf das gestellte Aufsatzthema konkret mit Zitaten aus dem Werk Bezug nehmen. In einigen Klassen dürfen zusätzlich auch Zusammenfassungen (résumé) oder eigene Notizen mitgebracht werden (siehe Tab. V.17).

Tab. V.17: Erlaubte Hilfsmittel (Maturitätsprüfungen Romandie) (Mehrfachnennungen möglich)

Matériel autorisé (Hilfsmittel)	F (N)	F (%)
Dictionnaires (Wörterbuch/-bücher)	20	37%
Œuvres (Schullektüre)	13	24%
Résumé (Zusammenfassungen)	5	9%
Notes (eigene Notizen)	2	4%
sans indication (ohne Angaben)	40	74%

Im Tessin durften alle untersuchten Klassen zwei Wörterbucher (I 100%) verwenden: sowohl ein gewöhnliches Wörterbuch (Dizionario della lingua italiana) als auch ein Synonym-Wörterbuch (Dizionario dei sinonimi e contrari).

Explizit ein PC erlaubt oder vorgeschrieben war nur in sechs der untersuchten Klassen der Deutschschweiz (5.7%). Im Tessin und in der Romandie fanden sich keine Hinweise auf den erlaubten Gebrauch von Computern.

3.1.2.10 Dauer der Prüfung

Keine Angabe zur vorgegebenen Zeit fand sich in rund der Hälfte der Fälle (D 49%, F 56%, I 0 %). In der anderen Hälfte hatten nahezu alle Klassen vier Stunden zur Verfügung (D 50%, F 44%, I 100%), nur drei Stunden Zeit erhielt eine Deutschklasse.

3.1.3 Kognitive Anforderungen

Einleitend kann festgehalten werden, dass alle Aufsätze der Maturitätsprüfung Erstsprache grundsätzlich dem höchsten der Hauptkategorie 6 entsprechen, also der Anforderung des Erschaffens (Elemente zusammensetzen, um ein kohärentes funktionales Ganzes zu formen; Elemente auf neuartige Weise strukturieren). Als Aufsätze erfordern sie im Speziellen die kognitive Leistung des „Erschaffens eines eigenständigen Produkts". Dies entspricht bei Anderson et al. (2001) der Kategorie 6.3.

Im Weiteren erfordert jeder Aufsatz eine oder mehrere zusätzliche Teilleistungen, die unterschiedlichen kognitiven Anspruchsniveaus nach Anderson et al. (2001) entsprechen. Diese Teilleistungen sind davon abhängig, welche Aufsatzform von der Maturandin oder dem Maturanden verlangt wird, welche Textsorte als Primärtext vorgelegt wird und wie die konkrete Aufgabenstellung formuliert ist. Jedes gestellte Thema erhält somit ein eigenes Profil, das auch mehrere Kategorien kognitiver Anforderungen aufweisen kann, falls die Lehrperson es in mehreren Teilaufgaben oder mehreren Arbeitsschritten formuliert hat. Codiert wurde dabei in der Erstsprache die explizite Aufgabenstellung, das heisst die ausformulierte Arbeitsanweisung der Lehrperson in vorgegebenen Teilschritten. Nicht erfasst wurden hingegen die impliziten kognitiven Teilleistungen, die bei der Bewältigung der verlangten Aufgabe zusätzlich von den Maturandinnen und Maturanden zu erbringen waren (beispielsweise bei der Interpretation einer Erzählung relevantes Wissen über die Autorin, die Epoche oder die Besonderheiten der Textsorte aus dem Langzeitgedächtnis abrufen).

Lag keine ausformulierte Arbeitsanweisung vor, was im Unterschied zur Deutschschweiz (8%) im Tessin beinahe hälftig (41%) und in der Romandie noch häufiger vorkam (67%), so wurde die Codierung des kognitiven Anspruchsniveaus aus der gegebenen Textsorte und ihren Themen sowie der verlangten Aufsatzform erschlossen.

Durch eine Gewichtung mit einem anteiligen Multiplikationsfaktor wurden Aufgabenstellungen, die mehrere Teilaufgaben enthielten, bei der Auszählung den Themen mit einer ganzheitlichen Aufgabenstellung wieder gleichgestellt: So erhielt bei einem Thema mit beispielsweise vier Teilaufgaben jede der Kategorien kognitiver Anforderungen nur einen Gewichtungsfaktor von 0.25, bei zwei Teilaufgaben je einen Faktor von 0.5, so dass jedes Thema mit einem Punkt gleich stark gewichtet ist.

Es wurde jeweils auf dem höchsten Anspruchsniveau im Sinne eines Anspruchspotenzials kodiert (zur Begründung siehe Abschnitt 1.2).

Durch eine Doppeleinstufung wurde dem oben beschriebenen Umstand Rechnung getragen, dass jeder Aufsatz grundsätzlich ein komplexes Produkt des Erschaffens ist (Kategorie 6.3), aber zusätzlich auch spezifische Teilleistungen gemäss dem jeweiligen Schwerpunkt verlangt: Der erste Teil der Doppelbenennung kennzeichnet die Kategorie des Erschaffens, die allen Aufsätzen gemeinsam ist. Der zweite Teil bezeichnet die Unterkategorie für die explizit verlangte Teilleistung innerhalb des jeweiligen Themas. Es ist zu beachten, dass jeder Maturaaufsatz implizit mehrere Kategorien kognitiver Anforderungen umfasst und der Fokus der Codierung auf die expliziten Teilleistungen aber für jede Aufgabe nur das dominierende kognitive Anspruchspotenzial berücksichtigen kann.

In der Deutschschweiz wurden 446 Aufgabenstellungen/Themen mit insgesamt 868 Teilaufgaben taxonomiert. Betrachtet man die Verteilung zusammengefasst auf die sechs Hauptniveaus kognitiver Anforderungen (siehe Tabelle V.18), so sind knapp ein Drittel aller schriftlichen Maturathemen der Deutschschweiz dem Anforderungsbereich des Erschaffen-Bewertens zuzuordnen. Bei allen diesen Aufgabenstellungen mussten die Maturandinnen und Maturanden einen Aufsatz verfassen, in welchem sie auf eigene Kriterien und zu erläuternde Standards gestützt ein Urteil abzugeben hatten. Auf der hier dominanten Kategorie 5.2 geht es darum, ein Phänomen oder eine Entwicklung kritisch zu untersuchen, mit Pro-Contra-Argumentationen Stellung zu beziehen und eine These, eine Haltung oder einen Prozess zu beurteilen. Meist handelte es sich hier um Erörterungen.

Tab. V.18: Analyse der kognitiven Anforderungen (Maturaprüfungen in der Deutschschweiz, Hauptkategorien)

Kognitive Hauptkategorien Deutsch (gewichtet)	Häufigkeit (D)
Erschaffen-Erinnern (1)	2.3%
Erschaffen-Verstehen (2)	19.1%
Erschaffen-Anwenden (3)	1.5%
Erschaffen-Analysieren (4)	22.2%
Erschaffen-Bewerten (5)	32.0%
Erschaffen-Erschaffen (6)	22.9%
Total Kognitive Hauptkategorien (gewichtet)	100.0%

Knapp ein Viertel aller Maturathemen entsprechen dem Bereich des Erschaffen-Analysierens. Dies sind Aufgabenstellungen, welche die Maturandinnen und Maturanden auffordern, vorgegebenes Material in seine Bestandteile zu zerlegen und zu bestimmen, wie die Teile untereinander in Beziehung stehen und wie sie mit der Gesamtstruktur zusammenhängen. Auf der hier dominanten Kategorie 4.3 geht es insbesondere darum, den Blickwinkel und die Absichten des Autors oder der Autorin zu erkennen und den Text zu interpretieren. Häufig handelt es sich um Interpretationen von Gedichten oder Kurzgeschichten. Knapp ein Viertel aller Maturathemen sind dem Bereich Erschaffen-Erschaffen zuzuordnen. Hier finden sich meist ziemlich komplexe Aufgabenstellungen, durch welche die Maturandinnen und Maturanden – oft auch ohne konkrete Arbeitsaufträge und ohne Textvorlage – gefordert sind, ein vollständig neues Produkt zu erfinden. Beispielsweise, indem sie eine eigene Kurzgeschichte entwerfen oder einen anspruchsvollen Essay gestalten. Knapp ein Fünftel aller Maturathemen gehört zum Bereich des Erschaffen-Verstehens. Hier sollen die Maturandinnen und Maturanden in der Lage sein, aus dem vorgegebenen Material Bedeutung zu gewinnen, indem sie allenfalls eigene Beispiele zur Veranschaulichung finden (Kategorie 2.2), Vergleiche ziehen (Kategorie 2.6) oder Ursache-Wirkungs-Modelle für ein Phänomen entwerfen (Kategorie 2.7).

Es ist auffallend, dass sowohl im Bereich Erschaffen-Erinnern, in dem es um das Abrufen von Wissen geht, sowie auch im Niveau 3, wo die Anwendung von Wissen gefragt ist, kaum explizite Aufgabenstellungen vorhanden sind. Damit kann festgestellt werden, dass der Fokus der schriftlichen Maturaprüfungen eindeutig nicht auf einer expliziten Wissensüberprüfung liegt. Wie stark jedoch das im Aufsatz dargelegte Wissen der Maturandinnen und Maturanden trotzdem in die Bewertung der jeweiligen Lehrperson einfliesst, kann mit den vorliegenden Daten nicht beantwortet werden.

Wie der Tabelle V.19 zu entnehmen ist, dominieren in der Deutschschweiz innerhalb der 19 Kategorien nach Anderson et al. (2001) eindeutig die drei Kategorien 5.2 Erschaffen-Beurteilen (D 30.9%), 4.3 Erschaffen-Erkennen (D 17.8%) und 6.3 Erschaffen-Konstruieren (D 19.5%).

Tab. V.19: Analyse der kognitiven Anforderungen (Haupt- und Unterkategorien, gewichtet[32]; Doppelbenennung) (Maturitätsprüfungen Deutschschweiz)

Kognitive Anforderungen Deutsch (gewichtet)		Gewichtungspunkte (D)	Häufigkeit (D)
Erinnern	1.1 Erschaffen-Identifizieren	2	0.4%
	1.2 Erschaffen-Abrufen	8	1.9%
Verstehen	2.1 Erschaffen-Interpretieren	2	0.5%
	2.2 Erschaffen-Veranschaulichen	30	6.6%
	2.3 Erschaffen-Klassifizieren	1	0.2%
	2.4 Erschaffen-Zusammenfassen	6	1.4%
	2.5 Erschaffen-Folgern	7	1.6%
	2.6 Erschaffen-Vergleichen	25	5.6%
	2.7 Erschaffen-Erklären	14	3.2%
Anwenden	3.1 Erschaffen-Ausführen	4	0.8%
	3.2 Erschaffen-Anwenden	3	0.7%
Analysieren	4.1 Erschaffen-Unterscheiden	4	0.8%
	4.2 Erschaffen-Integrieren	16	3.6%
	4.3 Erschaffen-Erkennen	79	17.8%
Bewerten	5.1 Erschaffen-Überprüfen	5	1.1%
	5.2 Erschaffen-Beurteilen	138	30.9%
Erschaffen	6.1 Erschaffen-Entwickeln	12	2.8%
	6.2 Erschaffen-Planen	3	0.6%
	6.3 Erschaffen-Konstruieren	87	19.5%
Total Gewichtungspunkte[33]		*446*	100.0%

Im Tessin fallen die Ergebnisse insofern ähnlich wie in der Deutschschweiz aus (siehe Tabelle V.20), als sich im Bereich des Erschaffen-Erinnerns (Kategorie 1) und Erschaffen-Anwendens (Kategorie 3) sehr wenige Aufgabenstellungen finden ([1] D 3.8%, I 6.1%), rund ein Fünftel der Aufgabenstellungen aus der Kategorie Erschaffen-Verstehen stammen (D 19.1%, I 19.2%), und der Grossteil der Prüfungsaufgaben dem oberen Teil der kognitiven Anforderungen (Kategorien 4–6) zugeordnet werden kann (D 77.2%, I 74.7%). Den Hauptteil stellen in der italienischsprachigen Schweiz jedoch die Aufgabenstellungen aus dem Bereich Erschaffen-Analysieren, der knapp ein Drittel aller Aufgabenstellungen ausmacht, gefolgt vom Bereich Erschaffen-Erschaffen mit einem Viertel aller Themen und „nur“ einem Fünftel aller Prüfungen innerhalb des Bereichs Erschaffen-Bewerten.

[32] Zum Verfahren der Gewichtung und der Doppelbenennung siehe vorhergehende Erläuterungen in 3.1.3

[33] Die Gewichtungspunkte entsprechen der Anzahl der Aufgabenstellungen aller Klassen.

Tab. V.20: Analyse der kognitiven Anforderungen (Maturitätsprüfungen im Tessin)

Kognitive Anforderungen Tessin (gewichtet)	Gewichtungspunkte (I)	Häufigkeit (I)	Gruppen
Erschaffen-Erinnern (1)	10	6,1%	
Erschaffen-Verstehen (2)	31	19,2%	
Erschaffen-Anwenden (3)	0	0,0%	25,3%
Erschaffen-Analysieren (4)	48	29,8%	
Erschaffen-Bewerten (5)	31	19,1%	
Erschaffen-Erschaffen (6)	42	25,8%	74,7%
Total Gewichtungspunkte	162	100,0%	100,0%

In der Romandie (siehe Tabelle V.21) fällt die Dominanz der sogenannten Beurteilungsaufgaben noch stärker ins Gewicht als in der Deutschschweiz; hier sind es zwei Drittel aller Aufgabenstellungen, die dem Bereich Erschaffen-Bewerten (Niveau 5) zugeordnet wurden. Es fällt auf, dass die Mehrzahl aller Themen und damit deutlich mehr (+14%) als in den beiden anderen Sprachregionen zur oberen Hälfte der Anforderungsbereiche (Niveau 4–6) gehört (F 91.3%).

Tab. V.21: Analyse der kognitiven Anforderungen (Maturitätsprüfungen in der Romandie)

Kognitive Hauptkategorien Romandie (gewichtet)	Häufigkeit (F)
Erschaffen-Erinnern (1)	0.2%
Erschaffen-Verstehen (2)	8.5%
Erschaffen-Anwenden (3)	0.0%
Erschaffen-Analysieren (4)	24.2%
Erschaffen-Bewerten (5)	66.1%
Erschaffen-Erschaffen (6)	1.0%
Total Kognitive Hauptkategorien (gewichtet[34])	100.0%

3.1.4 Vergleich mit den Aufgaben in Teilprojekt C

Im Unterschied zu den Maturitätsprüfungen wurde mit den Erstsprache-Tests von Teilprojekt C hauptsächlich das Leseverstehen untersucht; es kann damit etwas ausgesagt werden über die Fähigkeiten, welche die angehenden Studierenden aufweisen in den Aspekten allgemeines und detailliertes Leseverständnis sowie Wortschatz-Verständnis und grammatische und orthografische Kompetenzen (Sprachreflexionskompetenzen). Nur vereinzelt wurden in Test C auch schriftliche Produktionsaufgaben gestellt (eigene Titel setzen, Sätze umbilden), bei diesen handelte es sich aber immer um Kurzaufgaben. Somit können deren Ergebnisse nicht mit den Aufsatzformen der Maturaprüfungen verglichen werden.

In den Maturitätsprüfungen Erstsprache, die in Teilprojekt D1 untersucht wurden, liegt das Schwergewicht der Prüfung im Unterschied zu Teilprojekt C auf der schriftlichen Produktion. Das in Teilprojekt C getestete Leseverständnis ist dabei eine unabdingbare Voraussetzung, um überhaupt zum in Maturitätsprüfungen geforderten

[34] Zum Verfahren der Gewichtung und der Doppelbenennung siehe vorhergehende Erläuterungen in 3.1.3

Produkt zu gelangen: Es muss zuerst sowohl die Aufgabenstellung selber als auch der zitierte Primärtext richtig gelesen und verstanden werden. Erst auf dieser Grundlage können überhaupt eigene Überlegungen angestellt, Gedankengänge aufgebaut und ein schriftliches Produkt entwickelt werden.

Es gibt jedoch auch weitere Unterschiede zwischen den beiden Anforderungssituationen: Die Aussagen über das Leseverständnis in Teilprojekt C sind jeweils bezogen auf fachorientierte wissenschaftliche Sachtexte. Das bei den Maturaprüfungen geforderte Leseverstehen bezieht sich demgegenüber zu gut 40% bis 60% auf literarische Texte. Literarische Texte zeichnen sich dadurch aus, dass ihre Sinnerschliessung einen komplexen vielschichtigen Vorgang darstellt, bei dem Figurenkonstellationen, symbolische Verweise, Widersprüche, Verdichtungen und andere ästhetische Phänomene auf eine Weise zu erfassen sind, die nichts mit einer reinen Informationsentnahme zu tun hat. Dieses interpretierende Lesen ästhetischer Texte und das verstehende Lesen wissenschaftlicher Fachtexte bilden verschiedene Teilkompetenzen im Bereich Lesen ab. Auch die Sachtexte, die rund 20% der Maturaprüfungen zu Grunde liegen, entsprechen nicht derselben Textsorte wie die bei Teilprojekt C verwendeten Universitätstexte; zu einem grossen Teil handelt es sich bei den Sachtexten in Teilprojekt D1 um journalistische Texte oder politische Reden, nur sehr selten (4%) kommt auch wissenschaftliche Fachsprache vor.

Im Sprachreflexionsteil der Tests C wurden Sprachleistungen im Bereich Orthografie, Interpunktion, Grammatik und Wortschatz getestet. Diese Komponenten sind wichtige Bausteine für die Produktion schriftlicher Texte.

3.1.5 Vergleich mit den Ergebnissen aus Teilprojekt A4[35]

Ein Vergleich der Ergebnisse des Teilprojekts D1 (Einschätzung der Maturaaufsätze) mit den Ergebnissen aus Teilprojekt A4 (Befragung der Dozierenden) versucht die Frage zu beantworten, ob die schriftlichen Maturitätsprüfungen tatsächlich diejenigen Leistungen, Fähigkeiten und Kenntnisse überprüfen, welche die Dozierenden an Schweizerischen Hochschulen von den kommenden Studierenden erwarten. In einem weiteren Schritt geht es um die Frage, ob die von den Dozierenden bemängelten Wissens- und Könnens-Lücken der Erstsemestrigen sich bereits ansatzweise in der Überprüfung der Abschlussleistung abbilden.

Erstsprache hat als gymnasiales Grundlagenfach aus Sicht der befragten Dozierenden grundsätzlich die höchste Bedeutung: 60% aller Dozierenden erwarten fundiertes Wissen und Können der Gymnasiastinnen und Gymnasiasten in Erstsprache und 22% erwarten die Kenntnis wichtigster Grundlagen (Frage 1). Dies ist im Vergleich mit den Erwartungen gegenüber den anderen gymnasialen Fächern der höchste Wert. Eine detaillierte Betrachtung der konkreten Erwartungen, welche die Dozierenden an die vorausgesetzten Kenntnisse und Fähigkeiten in Erstsprache haben, führte zu folgenden Resultaten: Mit 38% steht die Sprachbeherrschung im Vordergrund der Erwartungen, gefolgt von der Beherrschung der Grammatik (10%), der Orthografie (4.3%,) und der Syntax (3.6%). Interpunktions-Kenntnisse werden explizit nur von 1% der Dozierenden erwartet. Spezifische Fachkenntnisse aus dem Fach Deutsch wie Literaturgeschichte werden nur von knapp 6% aller Dozierender erwartet, Analyse- und Interpretationsmethoden nur von knapp 3%, dabei stammen die meisten Nennungen fachspezifisch aus dem Feld der Germanistik-Dozierenden oder von anderen Sprach-Dozierenden.

Die erwartete Sprachbeherrschung auf hohem Niveau umschreiben einzelne Dozierende wie folgt näher: „Sprachbeherrschung auch in stilistischer Hinsicht“, „Sicherheit in Sprache und Schrift“, „schriftliche und mündliche Ausdrucksfähigkeit auf hohem Niveau“, „Fähigkeit zur präzisen sprachlich differenzierten Erfassung von historischen Inhalten“, „schriftlicher und mündlicher Ausdruck in einer Kombination aus Umgangs- und Fachsprache“, „stilsicherer Ausdruck“. Konkret zum Produzieren von Texten werden in Teilprojekt A4 folgende Erwartungen genannt: „wissenschaftliches Schreiben“, „Aufsatzgrundlagen“, „fehlerfreies Verfassen deutscher Texte“, „argumentative Texte verfassen können“, „präzises Formulieren“, „breiter Wortschatz“ , „Verfassen eigener Texte“, „Fähigkeit zur korrekten sprachlichen Beschreibung“, „strukturierte Texte verfassen“. Zum Lese-

[35] Eine vergleichende Auswertung mit den Ergebnissen aus Teilprojekt A1 und A3 kann erst nach weiterführenden Arbeiten erfolgen.

verständnis gab es folgende exemplarische Äusserungen: „exakte Textlektüre“, „Kenntnisse zur Verarbeitung von Fachliteratur“, „Fähigkeit, komplexe Sachverhalte zu erfassen und wiederzugeben“, „Sprachbeherrschung auch für komplexe Texte“.

Beim Vergleich dieser Erwartungen an das gymnasiale Fach Erstsprache mit den Maturitätsprüfungen lässt sich feststellen, dass die Art der Maturitätsprüfung Deutsch den Erwartungen der Dozierenden grundsätzlich entgegenkommt: Mit dem schriftlichen Produkt „Aufsatz“ zeigen die Maturandinnen und Maturanden neben der inhaltlichen Auseinandersetzung mit dem gesetzten Thema ihre sprachlichen Fähigkeiten ganz allgemein in den Bereichen „Sprachbeherrschung“, „schriftlicher Ausdruck“, „Stilsicherheit“ sowie ihre angewandten Grammatik-, Orthografie- und Interpunktionskenntnisse.

Gemäss der Analyse der geforderten Aufsatzformen (siehe Abschnitt 3.1.2.1) wird in der Deutschschweiz in 24% aller Aufsätze eine argumentative Textform (Erörterung, Argumentation) geprüft, in der Romandie sogar in 53% (dissertation, texte argumentatif). Dies entspricht ungefähr den Ergebnissen der Analyse der kognitiven Anforderungen (Kap. 3.1.3), gemäss denen im Tessin ein Fünftel, in der Deutschschweiz ein Drittel und in der Romandie zwei Drittel aller Fragestellungen im Anforderungsbereich Erschaffen-Bewerten liegen. Auf der meist dominanten Kategorie 5.2 müssen die Maturandinnen und Maturanden zeigen, dass sie fähig sind, mit Pro-Contra-Argumentationen Stellung zu beziehen und eine These zu beurteilen. Diese Ergebnisse des Teilprojekts D1 belegen, dass immerhin teilweise argumentative und damit vorwissenschaftliche Schreibformen im Sinne der Erwartungen von Dozierenden geprüft werden. Den Erwartungen der Dozierenden auf der Ebene des Leseverstehens kommt entgegen, dass beinahe allen Maturitätsprüfungen in Erstsprache ein oder mehrere Primärtexte zu Grunde liegen, deren exaktes Verständnis die Basis für den verlangten Aufsatz bildet.

Den Erwartungen der Sprach-Dozierenden als Subgruppe im Speziellen entspricht, dass die meisten Prüfungen Primärtexte aus dem Bereich „Literarische Texte“ (D 41%, F 62%, I 57%) enthalten und dabei eine Prüfung der erwarteten Kenntnisse von Interpretations- und Textanalysemethoden in der Anwendung erfolgt. Der von dieser Gruppe zusätzlich erwartete „Überblick über die wichtigsten Epochen der Literaturgeschichte“ wird jedoch kaum bis gar nicht geprüft.

Zu den Wissens- und Könnenslücken der neu eintretenden Studierenden befragt, wird hinsichtlich der Erstsprache Deutsch Folgendes bemängelt: Von einer insgesamt dürftigen Sprachbeherrschung sprechen 22 Dozierende, die ausschliesslich aus den nicht sprachlich orientierten Studienrichtungen stammen. Interessant dabei ist, dass 11 von 16 antwortenden Dozierenden der Rechtswissenschaften diese Lücke beobachten. Schlechte Beherrschung der Grammatik beklagen 14 Dozierende (vorwiegend aus Geschichte und Germanistik), fehlerhafte Rechtschreibung wird von 9 Dozierenden bemängelt, 4 Dozierende vermissen die Beherrschung der Syntax und der korrekten Interpunktion. Mangelhafte Kenntnisse der Literaturgeschichte mahnen 5 Dozierende der Germanistik und Anglistik an, fehlendes Wissen zur Interpretation von literarischen Texten 4 Dozierende (Germanistik, Anglistik), bessere Kenntnisse über Linguistik fordern zwei Sprach-Dozenten. Dass das Verfassen wissenschaftlicher Texte wie auch der Umgang mit wissenschaftlichen Texten Mühe macht, wird nur je einmal geäussert.

Ausgehend von diesen genannten Lücken bei den neu eintretenden Studierenden kann wiederum eine Übereinstimmung mit der Prüfungsart der Maturaprüfungen insofern festgestellt werden, als das entsprechende Wissen und Können eigentlich geprüft wird (Aufsatz als Ausdruck von Sprachbeherrschung, inklusive Grammatik, Syntax, Orthografie, Interpunktion). Es muss aber darauf hingewiesen werden, dass immerhin 18% der im Jahre 2007 vergebenen Noten in der schriftlichen Erstsprache-Prüfung an der Matura ungenügend waren. Solange die Kompensation ungenügender Noten durch genügende in anderen Fächern möglich ist, werden immer Lücken feststellbar sein.

Zusätzlich zu den oben erwähnten Fähigkeiten und Kenntnissen werden in den Maturitätsprüfungen Erstsprache implizit auch die weiteren überfachlichen Fähigkeiten „kritisches Denken“ und teilweise das „Entwickeln neuer Ideen“ geprüft (Fragen 4 und 5 in Teilprojekt A4). Das Erstere wird in den beschriebenen argumentativen Aufsatzformen geprüft, das Letztere wird bei Aufgabenstellungen der kognitiven Kategorien 6.1 „Entwickeln“

und 6.3 „Konstruieren“ erwartet. In der Deutschschweiz erfüllen dies 3% (Kategorie 6.1, Erschaffen-Entwickeln) bzw. sogar 20% aller Maturathemen (Kategorie 6.3, Erschaffen-Konstruieren).

Die in D1 ermittelten vielfältigen Prüfungsinhalte (Autorinnen und Autoren, Gesellschaftsthemen, ethische Probleme u.a.) stehen jedoch im Grossen und Ganzen nicht im Erwartungshorizont der befragten Dozierenden aller Fachrichtungen, dürften jedoch dem zweiten, in EVAMAR II nicht näher untersuchten Bildungsziel der schweizerischen Maturitätsausbildung, der Vorbereitung auf anspruchsvolle Aufgaben in der Gesellschaft (MAR, 1995, Art. 5), entsprechen.

Ergänzend sollen an dieser Stelle auch die speziellen Erwartungen der Germanistik-Dozierenden gemäss Teilprojekt A4 mit den in den schriftlichen Maturaprüfung Deutsch geprüften Kompetenzen gegenübergestellt werden. Es geht also nur noch um eine fachspezifische (statt allgemeine) Studierfähigkeit.

Die befragten Germanistik-Dozierenden haben folgendes Wissen und Können explizit als Anforderungen an zukünftige Germanistikstudierende formuliert (siehe auch Abschnitt 3 in Kapitel II):

- „erste Kenntnisse der Literaturgeschichte“ bis „Überblick über die wichtigsten Werke und Epochen der Literaturgeschichte“
- „Erfahrungen mit eigenständiger Textanalyse“ und „Grundkenntnis einiger Analysemethoden“ bis „Interpretationskompetenz“
- „Beherrschung grundlegender linguistischer und literaturwissenschaftlicher Begriffe“
- „Kenntnisse von Textsorten, Medien und Kommunikationsformen, Bewusstsein der Historizität von Sprache“
- „Kenntnis der Grammatik (Wortarten, Satzglieder, Morphologie, Syntax; Flexion von Substantiv und Adjektiv, Unterscheidung starker und schwacher Verben)“
- „Grundlagen der Sprachstruktur des Deutschen, grammatische Terminologie, Fähigkeit, sprachstrukturelle Regularitäten zu erkennen und zu formulieren“
- „Kenntnisse der wichtigsten Regeln abendländischer Poetik, lesetheoretische Grundmodelle in Wort und Tat“.

Wie bei anderen wissenschaftlichen Fächern wurden aber auch ein allgemein gutes Textverständnis, eine gute Schreibfähigkeit und eine „einwandfreie Kenntnis des Hochdeutschen“ genannt (siehe auch Abschnitt 3.1.5).

Es lässt sich feststellen, dass diese explizite fachspezifische Befähigung zu einem Germanistikstudium durch die oben in diesem Kapitel beschriebene schriftliche Maturitätsprüfung Deutsch *eher nicht* geprüft wird. Von den rund 40% der Aufgabenstellungen, die auf literarischen Primärtexten aus Schweizer und Deutscher Literatur beruhen, und sich somit im Bereich der Germanistik bewegen, sind einerseits 26% kurzzeilige Aphorismen, und andererseits stammen rund 62% aus dem 20. und 21. Jahrhundert. Damit bewegen sich diese Vorlagen eher am Rande des breiten Wissenschaftsbereichs der Germanistik bzw. nur in Teilbereichen. Insgesamt wurde bei der Analyse der schriftlichen Maturitätsprüfungen auch festgestellt, dass kaum Anforderungen an explizite Epochenkenntnisse, an deklaratives Wissen aus Gattungs-, Text- oder Grammatiktheorie oder an Grundlagenwissen der Literaturgeschichte sowie Literatur- und Sprachwissenschaft gestellt werden. Insofern prüfen die schriftlichen Maturitätsprüfungen nur sehr wenig explizites Vorwissen, das für ein Germanistikstudium fachspezifisch notwendig wäre bzw. fachspezifisch erwartet würde. Nicht untersucht wurden in EVAMAR II allerdings die mündlichen Deutsch-Prüfungen. Ein Urteil darüber, wie weit sich diese näher am Vorwissen und -können eines Germanistikstudiums bewegen, ist deshalb nicht möglich.

3.1.6 Gesamteinschätzung der Schweizer Maturaprüfungen in Erstsprache

Die schriftlichen Schweizer Maturaprüfungen in Erstsprache zeigen sich in ausserordentlicher Breite und Vielfalt: Es scheint, als ob jede Schule und jeder Kanton, ja jede Lehrperson, eigene Vorstellungen darüber hat, wie eine schriftliche Maturaprüfung auszugestalten ist und unter welchen Bedingungen die so genannte „Hochschulreife“ überprüft werden soll.

Die grösste Übereinstimmung hat die Schweizer Lehrerschaft darin, dass sie hauptsächlich Texte in der jeweiligen Originalsprache und selten Übersetzungen vorlegt, sowie darin, dass sie selten Texte mit weiblicher Autorschaft berücksichtigt. Ziemlich einig sind sich die Schweizer Lehrpersonen auch darin, dass sie meistens einen bis zwei Primärtexte aus dem 20. oder 21. Jahrhundert zu einem möglichst aktuellen oder zeitlosen Thema vorlegen. Dabei wählen sie zu rund 40 bis 60% literarische Primärtexte aus und nur zu 10 bis 20% Sachtexte; den Rest bilden so genannte „philosophische Texte“ (D 12%, F 14%, I 28%).

Vom Anspruchsniveau her gesehen, befinden sich drei Viertel bis die meisten der schriftlichen Maturitätsprüfungen (D 77%, F 91%, I 75%) im oberen Bereich der kognitiven Anforderungen (Niveau 4–6). Die grössten regionalen Unterschiede liegen in der eindeutigen Dominanz des Anforderungsbereiches „Erschaffen-Bewerten“ in der Romandie (zwei Drittel), im Unterschied zum dritten Rang dieses Bereiches im Tessin (mit einem Fünftel). In der Deutschschweiz ist die Verteilung ausgeglichener, indem zu einem Drittel Beurteilungsleistungen gefordert werden, zu einem Viertel Analysen und zu einem Viertel die Erfindung von Produkten. Insgesamt gehört die vierstündige Aufsatzform klar in die kognitive Kategorie des Erschaffens.

Inhaltlich zeugen die vorgelegten Themen von grosser Aktualität, der Orientierung an der Lebenswelt der Jugendlichen, einer starken Betonung der Meinungsvielfalt und von einer intensiven Auseinandersetzung sowohl mit wichtigen gesellschaftlichen Fragen als auch mit zeitlosen Sinnfragen des Menschseins.

3.2 Mathematik

Für die Analyse wurden insgesamt 168 schriftliche Maturaprüfungen untersucht: 101 aus der Deutschschweiz, 61 aus der französischsprachigen und 6 aus der italienischsprachigen Schweiz. Mit einigen wenigen Ausnahmen waren die Prüfungen in der entsprechenden Muttersprache verfasst. Es lagen auch Prüfungen für das Schwerpunktfach und für das Ergänzungsfach vor. Der Anteil der Schwerpunktfachprüfungen war in der Romandie grösser.

3.2.1 Das Analyse-Instrumentarium

Für die inhaltliche Analyse der Aufgaben wurde der in Teilprojekt B entwickelte Kompetenzraster verwendet.

Für die Analyse des kognitiven Aktivierungspotenzials der Aufgaben wurde das in Abschnitt 1.2 beschriebene Instrument für Mathematik, wie in der folgenden Tabelle V.22 dargestellt, angepasst.

Tab. V.22: Modell der kognitiven Aktivierungspotenziale für Mathematik

Kategorie	Kognitive Prozesse	Alternative Bezeichnungen	Definitionen
1	Erinnern relevantes Wissen aus dem Langzeitgedächtnis abrufen		
1.1	wieder erkennen	identifizieren	Wissen, das mit dem vorgegebenen Material konsistent ist, im Langzeitgedächtnis auffinden
1.2	abrufen	erinnern	relevantes Wissen aus dem Langzeitgedächtnis abrufen
2	Verstehen aus Lehrinhalten (mündlich, schriftlich und grafisch) Bedeutung gewinnen		
2.1	interpretieren	klären, verdeutlichen, umschreiben, umformulieren, anders darstellen, übersetzen, übertragen, umarbeiten	vorgegebenes Material von einer Darstellungsform in eine andere bringen
2.2	Beispiele machen, veranschaulichen, beispielhaft zeigen	illustrieren, veranschaulichen, bildlich darstellen	ein bestimmtes Beispiel oder eine bestimmte Abbildung finden, welche(s) ein Konzept oder ein Prinzip veranschaulicht
2.3	klassifizieren	kategorisieren, subsumieren	entscheiden, dass etwas zu einer Kategorie (Konzept oder Prinzip) gehört

Kategorie	Kognitive Prozesse	Alternative Bezeichnungen	Definitionen
2.4	zusammenfassen	abstrahieren, verallgemeinern	ein Hauptthema oder die wichtigsten Punkte zusammenfassen
2.5	folgern	schliessen, ableiten, extrapolieren, interpolieren, voraussagen, hochrechnen	logische Schlüsse ziehen
2.6	vergleichen	kontrastieren, gegenüberstellen, zuordnen, Ähnlichkeiten suchen	Lösungswege von Aufgaben vergleichen
2.7	erklären	Modelle und Mechanismen konstruieren	modellieren, bekannte mathematische Modelle einsetzen
3	Anwenden in einer gegebenen Situation eine Prozedur ausführen oder einsetzen		
3.1	ausführen	–	eine Prozedur auf ein bereits vertrautes Problem anwenden
3.2	implementieren	anwenden, Gebrauch machen von	eine Prozedur auf ein nicht bereits vertrautes Problem anwenden
4	Analysieren vorgegebenes Material in seine Bestandteile zerlegen und bestimmen, wie die Teile untereinander in Beziehung stehen und wie die Teile mit der Gesamtstruktur oder einem Gesamtzweck zusammenhängen		
4.1	differenzieren	unterscheiden, auseinanderhalten, fokussieren, auswählen	relevante (bzw. wichtige) Teile und irrelevante (bzw. unwichtige) Teile innerhalb des vorgegebenen Materials unterscheiden
4.2	organisieren	Kohärenz finden, integrieren, grob skizzieren, analysieren, strukturieren	bestimmen, wie Elemente innerhalb einer Struktur zusammenpassen oder zusammenspielen
4.3	auf etwas zurückführen	dekonstruieren	ein neues Problem auf ein bereits bekanntes Problem zurückführen
5	Bewerten auf Kriterien und Standards abgestützt ein Urteil abgeben		
5.1	überprüfen	ins Zusammenspiel bringen, Fehler detektieren, überprüfen, testen	Behauptungen verifizieren oder widerlegen
5.2	kritisch untersuchen	beurteilen	erkennen, ob eine Prozedur bei einem gegebenen Problem angemessen ist
6	Erschaffen Elemente zusammensetzen, um ein kohärentes funktionales Ganzes zu formen; Elemente auf neuartige Weise strukturieren		
6.1	hervorbringen, entwickeln	hypothetisieren	modellieren, neue mathematische Modelle entwickeln
6.2	planen	gestalten	eine Vorgehensweise entwickeln, um eine Aufgabe zu lösen
6.3	produzieren	konstruieren	eine Prozedur neu erfinden

3.2.2 Die Prüfungsinhalte in Mathematik

3.2.2.1 Vorgehen

Die Inhalte der Maturaprüfungen wurden in Anlehnung an den im Teilprojekt B entwickelten Kompetenzraster klassifiziert. Die mathematischen Inhalte sind dort in zwölf Teilgebiete unterteilt. Im Folgenden werden dafür die nachstehenden Abkürzungen verwendet:

AF: Analysis/Funktionen
AI: Analysis/Infinitesimalrechnung
AR: Analysis/Folgen und Reihen
GT: Geometrie/Trigonometrie
GG: Geometrie/elementare und analytische Geometrie
GV: Geometrie/Vektorrechnung und Vektorgeometrie
SW: Stochastik/Wahrscheinlichkeit
SK: Stochastik/Kombinatorik
SS: Stochastik/Statistik

XM: Vermischtes/Mengenlehre
XL: Vermischtes/Logik
XA: Vermischtes/elementare Algebra

Jeder Maturaaufgabe wurden in der Regel fünf Punkte zugeordnet, verteilt auf diejenigen Teilgebiete, die hauptsächlich in der Aufgabe angesprochen sind. Bei Aufgaben, deren Lösung eine markant höhere Punktzahl einbrachte als die übrigen Aufgaben, wurde die Punktzahl entsprechend erhöht. Bei etlichen Prüfungen sind aber die für die Lösung der Aufgabe vergebenen Punktzahlen nicht aufgeführt. Hier wurde davon ausgegangen, dass die zu erzielende Punktzahl für jede Aufgabe gleich ist.

Bei den meisten Aufgaben sind selbstverständlich weitere mathematische Teilgebiete angesprochen, es wurden aber nur solche mit Punkten bedacht, die einen Hauptteil einer Aufgabe ausmachen. Es ist auch zu beachten, dass diese Punktezuteilung nur aufgrund der Aufgabenstellung selber vorgenommen wurde und nicht aufgrund des Lösungsaufwandes für die Aufgabe. Aus Ressourcengründen war es für die codierenden Fachdidaktiker unmöglich, die Maturaaufgaben selbst zu lösen und so die Punkte für die einzelnen Teilgebiete genauer festzulegen. Die mit diesem Verfahren festgestellten Punkte wurden schliesslich für jedes Teilgebiet summiert und in Prozentzahlen umgerechnet, da nicht bei allen Prüfungen die gleiche Anzahl an (Teil-)Aufgaben vorlag. Damit ergab sich für jede Maturitätsprüfung eine Art Häufigkeitsverteilung für die einzelnen Teilgebiete des Kompetenzrasters (siehe technischer Bericht).

Für die bei der Zuordnung auftretenden Probleme fanden sich folgende Lösungen:

- Bei den sehr zahlreich vorkommenden Kurvendiskussionen wurden 2.5 Punkte dem Gebiet AF und 2.5 Punkte dem Gebiet AI zugeordnet. Bei etlichen Prüfungen war nicht klar ersichtlich, was die Maturandinnen und Maturanden zu tun hatten; das Vorgehen bei Kurvendiskussionen scheint oft vorher abgesprochen worden zu sein. Manchmal wurde auch auf ein Vorgehen gemäss einer Formelsammlung hingewiesen.
- Bei Extremalproblemen in der Differenzialrechnung wurden in der Regel fünf Kreditpunkte dem Gebiet AI zugeordnet, ausser wenn sich für das Finden der zur Diskussion stehenden Funktion andere Gebiete als sehr wichtig erwiesen. Bei Extremalproblemen in der Vektorrechnung (z. B. kürzester Abstand zwischen zwei Geraden) wurden die Punkte dem Teilgebiet GV zugeordnet, obwohl solche Aufgaben in der Regel auch mit Differenzialrechnung gelöst werden können.
- Für das Lösen von Gleichungen oder Gleichungssystemen wurden nur dann Punkte vergeben (für das Teilgebiet XA), wenn eine Teilaufgabe der Maturitätsprüfung ausschliesslich darin bestand, eine Gleichung oder ein Gleichungssystem zu lösen.
- Aufgaben, die Kreise, Kugeln, Kegel oder Zylinder beinhalteten, wurden dem Gebiet GG (analytische Geometrie) und nicht dem Gebiet GV zugeordnet.

3.2.2.2 Ergebnisse

Die Ergebnisse der Analyse sind in Tabelle V.23 zusammengefasst. Die Zahlen sind auf ganze Prozentzahlen gerundet. In der letzten Spalte XX sind mathematische Teilgebiete zusammengenommen, die sich nicht in den vorgegebenen Inhaltsraster einfügen liessen (Kryptologie, Modulrechnen, Rechnen mit komplexen Zahlen und Funktionen, Lineare Algebra, Lineare Programmierung, Computer-Programmierung, Physik, Chemie usw.).

Tab. V.23: Prozentuale Verteilung der Aufgaben auf die verschiedenen Inhaltsbereiche

	AF	AI	AR	GT	GG	GV	SW	SK	SS	XM	XL	XA	XX
Deutschschweiz	15%	28%	6%	3%	5%	16%	14%	6%	2%	0%	0%	2%	2%
Romandie	18%	25%	2%	1%	7%	7%	13%	3%	2%	0%	0%	2%	21%
Tessin	11%	19%	5%	0%	4%	10%	9%	0%	0%	0%	0%	0%	43%
Total	16%	26%	4%	2%	6%	13%	13%	5%	2%	0%	0%	2%	10%

Aus der Tabelle ist ersichtlich, dass national die meisten Aufgaben der analysierten Maturaprüfungen zum Gebiet der Infinitesimalrechnung (AI) gehören, gefolgt von Aufgaben zu Funktionen (AF). In den drei Sprachregionen zeigt sich ein teilweise unterschiedliches Bild: In der Deutschschweiz werden z. B. Aufgaben zur Vektorrechnung (GV) etwa gleich häufig gestellt wie Aufgaben zur Wahrscheinlichkeitsrechnung, während in der Romandie Aufgaben zur Vektorrechnung einen viel kleineren Umfang einnehmen. Der Anteil der Aufgaben, die sich nicht eindeutig in den Inhaltsraster einfügen liessen (XX), ist in der Romandie hoch und in Tessin sehr hoch. Dies lässt sich damit erklären, dass hier der Anteil von Maturitätsprüfungen für das Schwerpunktfach grösser war als in der Deutschschweiz. Es gibt auch leere Felder und damit Inhalte, die nicht geprüft wurden. So kommen Aufgaben zur Mengenlehre (XM) und Logik (XL) als eigenständige Maturateilaufgaben nicht (mehr) oder nur ganz vereinzelt vor. Eigenständige Aufgaben zur elementaren Algebra (XA) machen nur 2% aus. Das sind vor allem Aufgaben, bei denen Gleichungen oder Gleichungssysteme zu lösen sind. Auch Aufgaben zu Folgen und Reihen (AR), zur Trigonometrie (GT) und zur Statistik (SS) sind selten.

Tabelle V.24 fasst die Teilgebiete so zusammen wie in Teilprojekt C als Grundlage zur Generierung der Aufgaben.

Tab. V.24: Prozentuale Verteilung der Aufgaben auf die zusammengefassten Inhaltsbereiche

	Analysis	Geometrie	Stochastik	Vermischtes	XX
Deutschschweiz	49%	24%	22%	2%	2%
Romandie	45%	15%	18%	2%	21%
Tessin	35%	14%	9%	0%	43%
Total	46%	21%	20%	2%	10%

Es ist ersichtlich, dass die Aufgaben zur Analysis in der Deutschschweiz und in der Romandie fast die Hälfte aller Maturaaufgaben ausmachen. Der Anteil an Geometrieaufgaben ist in der Romandie und im Tessin erheblich geringer als in der Deutschschweiz. Die Differenz lässt sich vor allem durch die geringere Anzahl an Aufgaben zur Vektorgeometrie (GV) erklären. Elementare Geometrie und analytische Geometrie (GG) haben dagegen in der Romandie einen leicht höheren Stellenwert.

3.2.3 Kognitive Anforderungen

Bei der Bestimmung der kognitiven Anforderungen ist zu beachten, dass auch für Mathematik die in Abschnitt 1.2 beschriebenen Grenzen einer präzisen Bestimmung des kognitiven Anforderungsniveaus gelten, und nur das Potenzial der kognitiven Aktivierung bestimmbar war. Zudem war es aus organisatorischen und zeitlichen Gründen nicht möglich, zu den Aufgaben auch die Lösungen zu erhalten oder selbst zu erstellen. Auch das ging zulasten einer genaueren Kategorisierung. Die Festlegung des kognitiven Aktivierungspotenzials erfolgte aufgrund der Fragestellungen, also aufgrund der Aufgabenstellungen selber.

In einem ersten Schritt wurden die meist durch Verben explizit formulierten Aufforderungen („zeige“, „berechne“, „löse“, „bestimme die Funktionsvorschrift“) codiert. Fragestellungen wie „Wie gross ist die Wahrscheinlichkeit?“ oder „Wie lautet die Gleichung der Kugel?“ wurden umformuliert zu Aufgabenstellungen mit Verben: „Berechne die Wahrscheinlichkeit ...“ oder „Bestimme die Gleichung der Kugel“. Im zweiten Schritt erfolgte die Zuordnung zu den Kategorien.

Tabelle V.25 enthält eine vollständige Auflistung von in 162 codierten Maturaprüfungen aufgetretenen Aufforderungen (Deutschschweiz und Romandie).

Tab. V.25: Kognitive Anforderungen der Aufgabenstellungen

Nr.	Aufgabenstellung/Aufforderung	Kategorie nach Anderson et al. (2001)	Anteil in %
1	Kurvendiskussion (explizite Aufforderungen)	3.1	2
2	Kurvendiskussion (keine expliziten Aufforderungen)	3.1	<0.5
3	Definitionsbereich einer Funktion bestimmen	3.1	<0.5
4	Wertebereich bestimmen	3.1	<0.1
5	Funktionswerte berechnen	3.1	1.5
6	Wertetabelle erstellen	3.1	<0.5
7	Funktionsgleichung aufstellen	3.1	3
8	Graph skizzieren	2.1	3
9	Funktionsgleichungen – Graphen zuordnen	2.1	<0.1
10	Asymptoten bestimmen	3.1	<0.1
11	Asymptotisches Verhalten untersuchen (x gegen plus/minus unendlich)	3.1	<1
12	Nullstellen bestimmen	3.1	<1
13	Schnittpunkte zweier Graphen bestimmen	3.1	<1
14	Ableitung berechnen	3.1	1
15	Steigung/Steigungswinkel berechnen	3.1	<0.5
16	Monotonieverhalten abklären	3.1	<0.1
17	Krümmungsverhalten abklären	3.1	<0.5
18	Maxima/Minima suchen	3.1	<0.5
19	Wendepunkte suchen	3.1	<0.5
20	Tangentengleichung bestimmen	3.1	1.5
21	Normalengleichung bestimmen	3.1	<0.1
22	Winkel zwischen Funktionsgraphen bestimmen	3.1	<0.5
23	Winkel zwischen Tangenten/Normalen berechnen	3.1	<0.5
24	Stammfunktion bestimmen	3.1	<0.5
25	einfaches Integral berechnen	3.1	<0.5
26	anspruchsvolles Integral berechnen	3.1	<1
27	krummlinig berandete Fläche berechnen	3.1	<4
28	Fläche eines Dreiecks berechnen	3.1	<0.5
29	Länge eines Kurvenstückes berechnen	3.1	<0.5
30	Volumen eines Rotationskörpers bestimmen	3.1	<2
31	Oberfläche eines Rotationskörpers bestimmen	3.1	<0.5
32	einfache Extremalaufgabe lösen	4.3	<1
33	anspruchsvolle Extremalaufgabe (Modellierung erforderlich)	6.1	3
34	Werte von Parametern bestimmen	3.1	4
35	Koordinaten von speziellen Punkten bestimmen	3.1	<1
36	Sachverhalt beweisen	5.1	<0.1

Nr.	Aufgabenstellung/Aufforderung	Kategorie nach Anderson et al. (2001)	Anteil in %
37	Näherungsverfahren durchführen	3.1	<0.5
38	geometrischer Ort von Punkten mit gewissen Bedingungen bestimmen	3.1	<0.5
39	Beweisaufgabe	5.1	<0.5
40	Geschwindigkeit und Beschleunigung eines Punktes berechnen	3.1	<0.1
41	Taylorpolynom bestimmen	3.1	<0.5
42	Halbwertszeit bzw. Verdoppelungszeit berechnen	3.1	<0.5
43	Wachstums- oder Zerfallsfaktor berechnen	3.1	<0.5
44	Zinseszins berechnen	3.1	<0.5
45	Differenzialgleichung aufstellen	3.1	<0.5
46	einfache Differenzialgleichung lösen	3.1	<0.1
47	anspruchsvolle Differenzialgleichung lösen	3.1	<0.5
48	Phasendiagramm zeichnen	3.1	<0.1
49	Gleichgewichtspunkte bestimmen	3.1	<0.5
50	Trajektorien zeichnen	3.1	<0.1
51	quadratische Form diskutieren	3.1	<0.5
52	Kegelschnitt skizzieren	2.1	<0.5
53	Koordinaten von speziellen Punkten bei Kegelschnitten bestimmen	3.1	<0.5
54	Koordinaten von speziellen Punkten bei Funktionen zweier Variablen bestimmen	3.1	<0.1
55	Sattelpunkte bestimmen	3.1	<0.1
56	Extremalpunkte bei Funktionen zweier Variablen bestimmen	3.1	<0.1
57	Schnittpunkt Kurve–Fläche bestimmen	3.1	<0.1
58	Gleichung einer Tangentialebene bestimmen	3.1	<0.5
59	Winkel zwischen Kurve und Tangentialebene berechnen	3.1	<0.1
60	Flächenschwerpunkt bestimmen	3.1	<0.1
61	Folgenglieder berechnen	3.1	<2.5
62	explizite Darstellung einer Folge bestimmen	3.1	<0.5
63	rekursive Darstellung einer Folge bestimmen	3.1	<0.1
64	Teilsummen berechnen	3.1	<1
65	Konvergenz abklären	5.1	<0.1
66	Zahl der Folgeglieder mit Bedingung bestimmen	3.1	<0.5
67	Grenzwert berechnen	3.1	1
68	jährliche prozentuale Zu-/Abnahme berechnen	3.1	<0.5
69	Berechnung am rechtwinkligen Dreieck durchführen	3.1	<0.5
70	Berechnung am allgemeinen Dreieck durchführen	3.1	<0.1
71	Fläche einer geometrischen (geradlinig begrenzten) Figur berechnen	3.1	<0.5
72	Parameterdarstellung einer Geraden in R^2 oder R^3 bestimmen	3.1	<1.5
73	Parameterdarstellung einer Ebene bestimmen	3.1	<0.1
74	Koordinatengleichung einer Ebene bestimmen	3.1	2

Nr.	Aufgabenstellung/Aufforderung	Kategorie nach Anderson et al. (2001)	Anteil in %
75	Koordinatengleichung einer Geraden in R^2 bestimmen	3.1	<0.5
76	Richtungsvektor einer Geraden bestimmen	3.1	<0.1
77	gegenseitige Lage zweier Geraden bestimmen	3.1	<0.5
78	Schnittpunkt Gerade-Gerade berechnen	3.1	<0.5
79	Schnittpunkt Gerade-Ebene berechnen	3.1	<2
80	Koinzidenz Punkt-Gerade oder Punkt-Ebene nachweisen	5.1	<0.5
81	Fläche eines Dreiecks oder Parallelogramms berechnen	3.1	<1
82	Volumen von Tetraeder, Pyramide oder Spat berechnen	3.1	<1
83	Winkel zwischen Ebenen berechnen	3.1	<1
84	Winkel zwischen Geraden (bzw. Richtungsvektoren) berechnen	3.1	<1
85	Winkel zwischen Gerade und Ebene bestimmen	3.1	<1.5
86	Orthogonalität von Geraden nachweisen	5.1	<0.5
87	Orthogonalität Gerade-Ebene nachweisen	5.1	<0.5
88	Parallelität von Geraden nachweisen	5.1	<0.1
89	Koordinaten eines Punktes berechnen	3.1	<1.5
90	Abstand zweier Punkte berechnen	3.1	<1
91	Abstand zweier Geraden bestimmen	3.1	<0.5
92	Abstand Punkt-Gerade berechnen	3.1	0.5
93	Abstand Punkt-Ebene berechnen	3.1	<1.5
94	Mittelpunkt einer Strecke berechnen	3.1	<0.5
95	Parameter in Geradengleichung bestimmen	3.1	<0.5
96	Länge einer Strecke bestimmen	3.1	<0.5
97	Gleichung eines Kreises bestimmen	3.1	0.5
98	Fläche eines Kreises berechnen	3.1	<0.1
99	Gleichung (Mittelpunkt und/oder Radius) einer Kugel bestimmen	3.1	<2
100	Tangentialebene an Kugel bestimmen	3.1	<1
101	Koordinaten eines Punktes bestimmen, der gewisse Bedingungen erfüllt	3.1	6
102	Abstand zweier Punkte auf einer Kugel berechnen	3.1	<0.5
103	Polyeder zeichnen	2.1	<0.1
104	Besonderheit eines Vierecks nachweisen	5.1	<0.5
105	Verfahren beschreiben	2.1	<0.5
106	geometrischen Ort beschreiben	2.1	<0.1
107	geometrischen Sachverhalt beweisen	5.1	<0.1
108	Wahrscheinlichkeit bei einstufigem Zufallsversuch berechnen	3.1	2.5
109	Wahrscheinlichkeit bei mehrstufigem Zufallsversuch berechnen	3.1	7
110	bedingte Wahrscheinlichkeit berechnen	3.1	<1
111	Baumdiagramm zeichnen	2.1	<0.5
112	Anzahl Versuche bei mehrstufigem Zufallsversuch berechnen	3.1	2

Nr.	Aufgabenstellung/Aufforderung	Kategorie nach Anderson et al. (2001)	Anteil in %
113	stochastische Unabhängigkeit nachweisen	5.1	<0.1
114	kombinatorische Grundaufgabe lösen	3.1	<2
115	Anzahl Objekte bei kombinatorischer Aufgabe bestimmen	3.1	<0.5
116	Wahrscheinlichkeit bei kombinatorischer Grundaufgabe berechnen	3.1	3
117	statistische Grundaufgabe lösen	3.1	<0.5
118	Wahrscheinlichkeitsverteilung bestimmen	3.1	<0.5
119	Erwartungswert bestimmen	3.1	2
120	Wahrscheinlichkeit bei vorgegebener Verteilung bestimmen	3.1	<0.5
121	Varianz berechnen	3.1	<0.1
122	Standardabweichung berechnen	3.1	<0.5
123	Wahrscheinlichkeitsintervall berechnen	3.1	<0.1
124	statistischen Test wählen und durchführen	5.2	<0.5
125	(Null-)Hypothese formulieren	3.1	<0.5
126	Testgrösse bestimmen	3.1	<0.1
127	Verwerfungsbereich bestimmen	3.1	<0.5
128	Fehler- oder Irrtumswahrscheinlichkeit berechnen	3.1	<0.5
129	Hypothese prüfen	5.1	<0.5
130	Spielstrategie beurteilen	5.2	<0.1
131	(Spiel-)Strategie entwickeln	6.3	<0.1
132	Verfahren beschreiben	2.1	<0.1
133	Alternativen/alternatives Verfahren aufzeigen	6.2	<0.1
134	Elemente einer Menge aufzählen	3.1	<0.1
135	logische Implikationen mit Worten beschreiben	2.1	<0.1
136	Gleichung aufstellen	3.1	<1
137	Gleichung lösen	3.1	2
138	Ungleichung lösen	3.1	<0.1
139	Gleichungssystem aufstellen	3.1	<0.1
140	Gleichungssystem lösen	3.1	<0.1
141	Gleichung in der Menge der komplexen Zahlen lösen	3.1	<0.1
142	Prozentwert berechnen	3.1	<0.5
143	Standardaufgabe der linearen Programmierung lösen	3.1	<0.1
144	Anderes		<1.5

Sehr oft verbergen sich hinter einer Aufforderung weitere: Lautet eine Aufgabe z. B. „Bestimme die Fläche zwischen den Graphen der beiden Funktionen“, so müssen meist zuerst die Schnittpunkte der beiden Graphen ermittelt werden. In einer andern Aufgabe kann dann aber das Bestimmen der Schnittpunkte zweier Graphen explizit als Aufgabe, also als Aufforderung vorgegeben sein. Aus dem eben geschilderten Grund ist eine quantitative Anteilsbestimmung der vorkommenden Aufforderungen sehr ungenau. Es folgen deshalb nur einige qualitative Bemerkungen:

Die häufigste Frage an schriftlichen Maturitätsprüfungen lautet: „Wie gross ist die Wahrscheinlichkeit, dass ...?“. Häufig sind auch Aufgaben wie „Bestimme einen Punkt so, dass ...“ oder „Für welche Werte des Parameters a gilt ...“.

Eine typische Maturaaufgabe enthält im Mittel etwa vier bis fünf explizit formulierte Aufforderungen. Es zeigt sich, dass die meisten dieser Anforderungen dem kognitiven Prozess des Anwendens (Kategorien 3.1 und 3.2) zuzuordnen sind. Anwendungsaufgaben in diesem Sinne sind beispielsweise: „Bestimme die Funktionsgleichung ...“, „Bestimme die Koordinatengleichung ...“ oder „Berechne die Wahrscheinlichkeit ...“. Ob es sich dann um Kategorie 3.1 (eine Prozedur auf ein bereits vertrautes Problem anwenden) oder Kategorie 3.2 (eine Prozedur auf ein nicht bereits vertrautes Problem anwenden) handelt, lässt sich meist nicht mit Bestimmtheit sagen (vergleiche dazu die Bemerkungen in Abschnitt 1.2). Die Autoren tendieren eher zur Einstufung in 3.1.

In jeder Maturaaufgabe in Mathematik wird relevantes Wissen vorausgesetzt. So enthält jede Aufgabe auch einen mehr oder weniger grossen Anteil der Kategorie „Erinnern“.

Sehr häufig sind auch Aufgaben, bei denen man zu vorgegebenen Funktionsgleichungen die entsprechenden Funktionsgraphen zeichnen oder einen mehrstufigen Zufallsversuch mittels eines Baumdiagramms darstellen muss. Diese Aufgabenstellungen können dem Bereich „Verstehen“ zugeordnet werden, entweder 2.1 oder 2.2.

Bei den (immer noch) beliebten Kurvendiskussionen wurde unterschieden, ob konkret angegeben war, was zu tun ist („Bestimme den Definitionsbereich, finde Nullstellen und Extrema ...“) oder ob sich die Maturandinnen und Maturanden zuerst daran erinnern mussten, was zu tun ist. Aufgaben mit Kurvendiskussionen wurden ebenfalls der Kategorie 3.1 zugeordnet.

In praktisch jeder Maturaprüfung in Mathematik kommt eine Extremalaufgabe vor. Einfachere Extremalaufgaben wurden der Kategorie 4.3 zugeordnet. Anspruchsvolle Extremalaufgaben, bei denen zuerst eine Übersetzung/Modellierung gefordert ist, können aber auch als 6.1 klassifiziert werden.

Beweise, wie sie besonders bei Prüfungen im Schwerpunktfach „Physik und Anwendungen der Mathematik“ (PAM) zu finden sind, gehören in die Kategorie 5.1. In die gleiche Kategorie wurden auch Aufforderungen wie „Weise nach, dass die Geraden windschief sind ...“, „Zeige, dass die Gerade g in der Ebene E liegt ...“ usw. eingeteilt. Obwohl man Aufgaben wie „Zeige, dass die Geraden sich im Raum schneiden“ natürlich auch lösen kann, indem man den Schnittpunkt berechnet (Kategorie 3.1), wurden sie der Kategorie 5 zugeordnet.

Die sehr seltenen Aufforderungen „Wähle einen geeigneten (statistischen) Test und führe diesen durch ...“ wurden der Kategorie 5.2 zugeordnet.

Im Zusammenhang mit Wahrscheinlichkeitsrechnung taucht ab und zu die Aufgabe auf, eine eigene Spielstrategie zu entwickeln. Solche Aufgaben wurden in die Kategorie 6.3 eingeteilt.

Die nachfolgende Tabelle V.26 enthält eine grobe quantitative Auflistung der in den untersuchten Maturaprüfungen gestellten kognitiven Anforderungen bzw. der kognitiven Anforderungspotenziale:

Tab. V.26: Quantitative Darstellung des kognitiven Aktivierungspotenzials

Kategorie	Anteil
2.1 Verstehen: interpretieren	< 5%
3.1 Anwenden: ausführen	85%
4.3 Analysieren: auf etwas zurückführen	< 1%
5.1 und 5.2 Bewerten: überprüfen und kritisch untersuchen	< 5%
6.1 und 6.3 Erschaffen: hervorbringen, entwickeln und produzieren	< 5%

(< 5% bedeutet, dass der prozentuale Anteil leicht unter 5% liegt.)

3.2.4 Vergleich mit den Aufgaben in Teilprojekt C

Die Testaufgaben in Teilprojekt C sind mit den Maturaaufgaben nur bedingt vergleichbar, denn meistens erfordern Maturateilaufgaben einen erheblich grösseren Lösungsaufwand, als dies bei Testaufgaben vorgesehen und möglich war. Ferner dürfen an Maturitätsprüfungen meistens Taschenrechner eingesetzt werden, was beim Mathematiktest nicht erlaubt war. Trotzdem lassen sich folgende Gemeinsamkeiten und Unterschiede finden:

Aufforderungen wie „Funktionswerte berechnen ...“, „Graphen skizzieren ...“, „Funktionsgleichungen – Graphen zuordnen ...“, „Ableitungen von Funktionen berechnen ...“, „Stammfunktion bestimmen ...“, „einfache Wahrscheinlichkeiten berechnen ...“ lassen sich sowohl bei den Testaufgaben wie auch in Maturaprüfungen finden. Eine Korrelation zwischen Lösungshäufigkeit beim Test und Häufigkeit in schriftlichen Maturaaufgaben lässt sich allerdings nicht ausmachen.

Testaufgaben wie „Graphen skizzieren“ wurden schlecht gelöst, offenbar, weil die vorgegebene Funktionsvorschrift nicht einfach zu verstehen war. Aufgaben zur Stammfunktion wurden in den Testaufgaben nur mittelmässig gelöst, obwohl fast in jeder Maturaprüfung eine Stammfunktion gesucht werden muss, um beispielsweise eine krummlinig berandete Fläche zu berechnen. Andere Testaufgaben wurden gut gelöst, obwohl solche Aufgaben fast nie in Maturaprüfungen auftauchen.

Allgemein lässt sich Folgendes festhalten: Aufgaben zur Analysis machen etwa 50% aller Maturaaufgaben aus. Trotzdem wurden diese Testaufgaben am schlechtesten gelöst. Der Unterschied in der Lösungshäufigkeit bei den Geometrieaufgaben zwischen der Deutschschweiz und der Romandie könnte daraus resultieren, dass Geometrieaufgaben in der Romandie weniger häufig gestellt werden. Testaufgaben zu Stoffgebieten, die an Maturitätsprüfungen nicht vorkommen (Mengenlehre, Logik) wurden erstaunlich gut gelöst. Ebenso eher ungewohnte Aufgaben zu geometrischen Abbildungen und Funktionen mit zwei Variablen. Maturaufgaben mit diesen Inhalten sind sehr selten und tauchen eigentlich nur in Prüfungen des Schwerpunktfachs Physik und Anwendungen der Mathematik auf.

3.2.5 Vergleich mit den Ergebnissen aus Teilprojekt A4

Tabelle V.27 enthält einen Vergleich der Ergebnisse der Befragung der Dozierenden in Teilprojekt A4 (siehe Abschnitt 3 in Kapitel II). Die erste Spalte listet Antworten auf die Frage nach bedeutsamen Wissensinhalten für die Lehrveranstaltungen auf (Frage 2). In der zweiten Spalte sind festgestellte Lücken aufgeführt (Frage 3). Die dritte Spalte enthält Kommentare und Vergleiche der von den Dozierenden gemachten Aussagen zu den Maturaprüfungen.

Tab. V.27: Vergleich der Anforderungen Dozierender mit den Maturitätsprüfungen

Konkrete Wissensinhalte aus den für die Lehrveranstaltung bedeutsamen gymnasialen Fachbereichen (Frage 2)	Aktuell festgestellte Wissens- und Könnenslücken (Frage 3)	Kommentar/Vergleich mit Aufgaben in Maturitätsprüfungen
Allgemeines	Allgemeines	
mathematische Sachverhalte ausdrücken können	Skizzen von mathematischen Sachverhalten	zu vage
wissenschaftstheoretische Grundlagen		zu vage
Beweiskonzept		zu vage
	flexibles anwendbares Wissen	zu vage
	allgemeine mathematische Kenntnisse	zu vage

Konkrete Wissensinhalte aus den für die Lehrveranstaltung bedeutsamen gymnasialen Fachbereichen (Frage 2)	Aktuell festgestellte Wissens- und Könnenslücken (Frage 3)	Kommentar/Vergleich mit Aufgaben in Maturitätsprüfungen
	mündlicher und schriftlicher Ausdruck	Ob der schriftliche Ausdruck bei Maturitätsprüfungen bewertet wird, lässt sich hier nicht feststellen.
Logik	Logik	Kommt in Maturitätsprüfungen praktisch nicht vor.
Algebra-Grundkenntnisse	Algebra	
mathematische Formeln	Formeln lesen	Wird vor allem im Zusammenhang mit Funktionsgleichungen und beim Beschreiben mathematischer Objekte (Geraden, Ebenen) mit Vektoren geprüft.
binomischer Lehrsatz		Kommt in den schriftlichen Prüfungen nicht vor.
Formelsprache, z.B. Summenzeichen		Eine angemessene Formelsprache ist in praktisch jeder Maturitätsprüfung vorhanden.
Umgang mit Variablen		Der Umgang mit Variablen ist für das Lösen der meisten Maturaaufgaben unerlässlich. Es gibt sehr viele Maturaaufgaben, bei denen Parameter so bestimmt werden müssen, dass gewisse Bedingungen erfüllt sind.
lineare Gleichungen	lineare und quadratische Gleichungen	Explizite Aufforderungen zum Lösen von Gleichungen oder Gleichungssystemen sind eher selten. Bei Aufgaben zur Vektorrechnung und zu Wachstums- und Zerfallsprozessen treten Gleichungen/Gleichungssysteme oft implizit auf.
Lösen von 1–2 Gleichungen mit 2 Unbekannten	Gleichungen mit 2 Unbekannten	siehe Kommentar „lineare Gleichungen“
Potenzgesetze	Potenzgesetze	Kommen bei schriftlichen Aufgaben nicht vor.
Funktionslehre	Funktionslehre	
allgemeiner Funktionsbegriff, Darstellung von Tabellen und Formeln		Der Umgang mit Funktionen macht gut die Hälfte aller Maturaaufgaben aus. Selten in Maturaprüfungen sind explizite Aufforderungen, Wertetabellen zu erstellen. Für das Skizzieren der Funktionsgraphen sind diese aber oft unerlässlich, ausser man beschränkt sich dabei auf spezielle Punkte (Nullstellen, Extrema, Wendepunkte).
Proportionalität	Proportionalität	Das Prinzip der Proportionalität taucht in Maturaprüfungen am häufigsten bei Prozentrechnungen auf. Sonst geht es meistens um die Beschreibung von Phänomenen, die sich eben gerade nicht proportional verhalten (exponentielle Wachstums- und Zerfallsprozesse).
Graphen elementarer Funktionen		Mehr als 90% aller Maturaprüfungen enthalten eine Aufforderung zum Zeichne/Skizzieren eines Funktionsgraphen.
Kurvendiskussion		Rund zwei Drittel aller Maturaprüfungen enthalten eine Aufforderung zur Kurvendiskussion.
trigonometrische Funktionen	trigonometrische Funktionen (Winkelfunktionen)	Kommen in Maturaufgaben nur noch sehr selten vor. Gehört in vielen Kantonen wohl nicht mehr zum Stoff der letzten beiden Ausbildungsjahre.
Exponential- und Logarithmusfunktionen	Exponential- und Logarithmusfunktionen	Bei der Analyse der Maturaprüfungen wurde nicht unterschieden, für welchen Typ von Funktionen zum Beispiel die Funktionsgleichung bestimmt werden muss. Aufgaben zu Wachstums- oder Zerfallsprozessen sind aber relativ selten.
Minimum und Maximum einer Funktion		Es gibt kaum eine Maturaprüfung, bei der nicht mindestens einmal ein Minimum oder ein Maximum einer Funktion gesucht werden muss.

Konkrete Wissensinhalte aus den für die Lehrveranstaltung bedeutsamen gymnasialen Fachbereichen (Frage 2)	Aktuell festgestellte Wissens- und Könnenslücken (Frage 3)	Kommentar/Vergleich mit Aufgaben in Maturitätsprüfungen
(analytische) Geometrie	(analytische) Geometrie/ Vektorrechnung	
Planimetrie (Spiegelung an Geraden, Drehungen)	Planimetrie (Bogenlänge und -mass)	Die analytische Beschreibung von geometrischen Abbildungen kommt nur ab und zu in Prüfungen des Schwerpunktfaches PAM vor. Aufgaben zu trigonometrischen Funktionen (bei denen zwingend im Bogenmass gerechnet werden muss) sind eher selten.
Strahlensatz		Kommt explizit in Maturaprüfungen nicht vor.
Koordinatensysteme		Koordinatensysteme sind Grundlage für das Darstellen von Funktionsgraphen. Ohne Verständnis des Koordinatensystems wird das Skizzieren von Graphen nicht gelingen.
Vektoren	Vektor (Verbindungsvektor, Länge des Vektors)	Praktisch jede Maturitätsprüfung enthält eine Aufgabe zur Vektorrechnung.
Vektor-Analysis		siehe oben
	analytische Geometrie	zu vage
Integral- und Differenzialrechnung	Differenzial- und Integralrechnung	
Differenzialrechnung: Ableitungen	Differenzialrechnung: Ableitungen	Praktisch jede Maturitätsprüfung enthält eine Aufgabe zu Ableitungen.
Physikalische Deutungen der Ableitung		Kommt nur in Aufgaben des Schwerpunktfaches PAM vor.
Integralrechnung		Praktisch jede Maturitätsprüfung enthält eine Aufgabe zur Integralrechnung
Mengenlehre		Kommt in Maturaufgaben nicht vor.
Stochastik	Stochastik	
Wahrscheinlichkeitsrechnung	Wahrscheinlichkeitsrechnung	Praktisch jede Maturitätsprüfung enthält eine Aufgabe zur Wahrscheinlichkeitsrechnung.
Statistik-Grundlagen	Statistik	Aufgaben zur Statistik sind eher selten.
Unterscheidung von Korrelations- und Kausalzusammenhang		Kommt so in Maturaprüfungen nicht vor.
Folgen und Reihen	Folgen/Reihen	
Definitionen für geometrische Folge und geometrische Reihe	Definition für geometrische Folge/Reihe	Nur etwa jede dritte Maturaprüfung enthält eine Aufgabe zu Folgen und Reihen. Häufig geht es dann um geometrische Folgen und Reihen, besonders häufig sind die Folgeglieder als geometrische Streckenlängen oder Flächeninhalte geometrischer Figuren gedeutet.
Generelle Bemerkungen		
„Normalerweise ist das Wissen in Linearer Algebra schwach“		Lineare Algebra ist nicht Stoff des Grundlagenfaches Mathematik.
starke individuelle Unterschiede		zu vage
„Gefühle für Grössenordnungen und Mühe mit abstraktem und analytischem Denken in einfachen Zusammenhängen“		Zu vage. Aufgaben, bei denen nur die Grössenordnung von Ergebnissen angegeben werden muss, kommen nicht vor.

Einige der als lückenhaft bemängelten Gebiete gehören offenbar nicht zu den in den Maturitätsprüfungen thematisierten Inhalten des Gymnasiums (Lineare Algebra, Logik usw.). Die Stoffgebiete Analysis, Vektorrechnung oder Stochastik sind in den untersuchten Maturaprüfungen reichlich vertreten, das heisst in praktisch allen Prüfungen. Aufgaben zur Trigonometrie (sowohl klassische Aufgaben wie auch solche zu trigonometrischen Funktionen und ihren Graphen) sind sehr selten, da Trigonometrie offenbar nicht zum Stoff der beiden letzten Ausbildungsjahre am Gymnasium bzw. der Maturaprüfungen gehört. Im Kanton Bern zum Beispiel wird an der Maturitätsprüfung schwergewichtig der Stoff der beiden letzten Jahre geprüft.

3.2.6 Gesamteinschätzung der Schweizer Maturaprüfungen in Mathematik

Inhaltlich liegt eine grosse Vielfalt vor. Die Maturaaufgaben zur Mathematik sind in aller Regel sorgfältig konstruiert und sehr phantasievoll ausgearbeitet.

Es gibt daneben aber auch eine stattliche Zahl von „Standardprüfungen“: Je eine Aufgabe zu Kurvendiskussion, zu Vektorrechnung, zu Wahrscheinlichkeitsrechnung, Differenzialrechnung (Extremalproblem) und eine Aufgabe mit mehreren Kurzaufgaben, mit der weitere Gebiete (Folgen und Reihen, elementare Algebra usw.) abgedeckt werden. Bezüglich des kognitiven Aktivierungspotenzials ist hier der Anteil an „Erinnern“ besonders hoch.

Grosse Unterschiede gibt es bei den (potenziellen) kognitiven Anforderungen, sowohl zwischen Grundlagenfachprüfungen und Schwerpunktfachprüfungen wie auch zwischen den Grundlagenfachprüfungen verschiedener Schulen. Das lässt sich aber erst im Detail spezifizieren, wenn man einzelne Prüfungen genauer untersucht und die Lösungen der Aufgaben vergleicht. Wie schon erwähnt, war das aus Zeitgründen nicht möglich.

Hinsichtlich des kognitiven Aktivierungspotenzials ist der Grossteil der Matura(teil-)aufgaben der Kategorie 3.1 (Anwenden einer Prozedur auf ein bereits vertrautes Problem) zuzuordnen. Die fast in jeder Prüfung vorkommende Extremalaufgabe gehört nach Einschätzung der Autoren entweder in die Kategorie 4.3 (ein neues Problem auf ein bereits bekanntes Problem zurückführen) oder, falls es sich um anspruchsvolle Extremalprobleme handelt, die ein hohes Mass an Modellierungsarbeit erfordern, in die Hauptkategorie 6 (Erschaffen).

3.3 Biologie

Für Biologie wurden Aufgaben aus 65 Maturitätsprüfungen des Jahres 2007 untersucht. Diese kleine Zahl ist darauf zurückzuführen, dass Biologie als Bestandteil des Grundlagenfachs Naturwissenschaften an den allermeisten Schulen nicht schriftlich geprüft wird. Von den untersuchten Prüfungen stammen die meisten aus der Maturaprüfung für das Schwerpunkt- oder das Ergänzungsfach. Nachfolgend wird zunächst das Analyseinstrumentarium beschrieben.

3.3.1 Das Analyseinstrumentarium

Sämtliche Aufgaben der Maturitätsprüfungen wurden inhaltlich gemäss dem Kompetenzraster Biologie (siehe Berichtteil III, Kap. 3) in sechs Teilbereiche kategorisiert, die in Tabelle V.28 dargestellt werden.

Tab. V.28: Inhaltsbereiche für die Analyse der Maturaaufgaben

Inhaltsbereich (angelehnt an deutschsprachige Standardbiologielehrwerke SII)					
Cytologie/Anatomie/ Stoffwechsel	Informations-verarbeitung/ Verhalten	Immunbiologie	Genetik/Entwicklungs-biologie	Ökologie	Systematik/Evolution

Die bewusst relativ grobe Unterteilung in nur sechs Teilbereiche zahlte sich bei der Kategorisierung aus: Die Aufgaben liessen sich in den allermeisten Fällen eindeutig zuordnen. Oft konnte man auch den Zwischentiteln folgen, welche die Autoren der Prüfungen bereits selber über die Aufgabengruppen gesetzt hatten.

Das Instrument zur Analyse des kognitiven Aktivierungspotenzials entspricht dem in Abschnitt 1.2 beschriebenen (siehe dort Tabelle V.1). Es ist in den Naturwissenschaften breit bekannt und gut akzeptiert. Die ausführliche Charakterisierung der Kategorien mit vielen Beispielen konkreter Aufgabenstellungen, die als Grundlage der Codierungsarbeiten diente, wird im technischen Bericht publiziert.

Für die Analyse mussten die Prüfungen weiter strukturiert werden. Im trivialen Fall ist vor allem in den Deutschschweizer Biologieprüfungen die maximal erreichbare Punktzahl für jede Aufgabe und Teilaufgabe direkt vom Prüfungsautor angegeben (in 58 von 60 Prüfungen). Von den insgesamt 10 untersuchten Prüfungen aus der Romandie ist dies nur für deren 4 der Fall. Wenn nur die Gesamtpunktzahl einer Aufgabe angegeben war, nicht aber die Aufteilung der Punkte auf die Teilaufgaben, wurde die Gesamtpunktzahl durch die Anzahl der Teilaufgaben dividiert. Dieses Vorgehen kommt ohne jede weitere Annahme aus, wenn die tatsächliche Bewertungsweise des Aufgabenautors nicht bekannt war.

Auch die Kategorisierung der Biologieaufgaben erfolgte ausschliesslich auf der Basis schriftlicher Originalprüfungen. Die spezifische in der Aufgabenstellung verwendete Wortwahl und Formulierung (vor allem der Verben) wurde bewusst nicht als Grundlage einer vermeintlich sicheren bzw. vollständig objektivierbaren Kategorieneinteilung gewählt. Die Kategorisierung stützt sich vielmehr auf die Einschätzung ab, welche kognitiven Prozesse beim Aufgabenlösen mit grosser Wahrscheinlichkeit aktiviert werden. In den meisten Fällen kann dies aufgrund von Plausibilitätsüberlegungen relativ gut abgeschätzt werden. Eine Aufgabe wie „Erkläre, wie die DNA mithilfe von Enzymen repliziert wird“ erfordert selbstredend von keinem Biologieschüler, dass er an der Prüfung auf der Grundlage von gymnasialen Kenntnissen der Biochemie, Enzymkinetik und der chemischen Struktur der DNA einen plausiblen (1959 tatsächlich mit dem Nobelpreis ausgezeichneten) Mechanismus der Replikation entwirft. Mit der allergrössten Wahrscheinlichkeit ist mit „Erklären“ im Sinne von „beschreiben“ ein einfacher Zugriff auf das Langzeitgedächtnis gemeint (Kategorie 1.2). Auf der anderen Seite erfordern einige Aufgaben eine Einschätzung, ob eine in der Hinführung zur Aufgabenstellung verwendete Darstellung einer Tatsache, einer Struktur, eines Vorgangs oder eines Experiments für eine Maturandin oder einen Maturanden bereits in sehr ähnlicher Art im Unterricht präsentiert worden ist oder ob es sich mit grosser Wahrscheinlichkeit um neuartiges Material handelt, das vom Testautor eigens für die Aufgabe aufbereitet wurde. Die vorliegende Kategorisierung verwendet als Haupterkennungsmerkmal für neuartiges Material, dass es analog einem kleinen Ausschnitt aus einem Lehrbuch ausführlich und in sich geschlossen vorgestellt wird, das heisst mit einem hohen Mass an Textkohäsion. Im Gegensatz dazu weist eine knappe Hinführung zur Aufgabe mit einigen isolierten und nicht weiter erläuterten Stichworten darauf hin, dass die Maturandin bzw. der Maturand den restlichen Zusammenhang aufgrund des vorangegangenen Unterrichts rekonstruieren kann.

Grundsätzlich galt als Haupterkennungsmerkmal für Aufgaben mit einem hohen Transferanteil,

- dass eine Frage eher erst am Schluss eines Blocks von Teilaufgaben steht, und zwar nach einigen Fragen der Kategorie 1.2 (Zugriff auf Langzeitgedächtnis);
- dass der Konjunktiv verwendet wird: „Wie würde das Resultat wohl aussehen?“, „Wieso wäre es möglich, dass ...?“.

3.3.2 Prüfungsinhalte in Biologie

Es soll untersucht werden, welche Abweichungen die inhaltliche Gewichtung der Aufgaben in schweizerischen Maturitätsprüfungen 2007 von einer allgemein akzeptierten „neutralen“ Gewichtung zeigt. Eine solche allgemein akzeptierte Gewichtung wird aus drei deutschsprachigen Standardlehrwerken der Biologie auf der Gymnasialstufe (Stufe SII) abgeleitet, und zwar anhand der Seitenzahl, welche die Themengebiete jeweils beanspruchen (siehe Tab. III.1) Um das relative Volumen abzuschätzen, das die sechs Themenbereiche des Kompetenzrasters

(siehe Abschnitt 4.3 in Kapitel III) in drei Standard-Biologielehrwerken je einnehmen, wurden die entsprechenden Seiten gezählt, und anschliessend wurde das mittlere relative Volumen des Teilgebiets bestimmt.

Es fällt auf, dass sich die relativen Kapitelvolumina in den drei untersuchten Standard-Biologielehrwerken sehr ähnlich sind (siehe die ausführlichen Tabellen im technischen Bericht). Dies spricht für die Plausibilität des Ansatzes, diese Prozentzahlen als Referenz im Sinne einer allgemein akzeptierten Gewichtung einzusetzen.

Der durchschnittliche Anteil des Teilbereichs *t* an der Gesamtpunktzahl einer Maturitätsprüfung 2007 wurde durch folgende Berechnung geschätzt:

- geg. Maturitätsprüfung *j*
- geg. Prüfungsteilfrage Q_{ij} = Prüfungsfrage *i* innerhalb Maturitätsprüfung *j*
- geg. n_G = Gesamtzahl der analysierten Maturitätsprüfungen = 65
- geg. z_j = maximal erreichbare Punktzahl für das Beantworten der ganzen Maturitätsprüfung *j*
- geg. a_{ij} = maximal erreichbare Punktzahl für das Beantworten der Prüfungsfrage Q_{ij}
- M_t = Menge aller a_{ij}, so dass die entsprechenden Prüfungsfragen Q_{ij} zu Teilbereich *t* gehören (gemäss der Inhaltsanalyse)
- ges. durchschnittlicher Anteil des Teilbereichs *t* an der Gesamtpunktzahl einer Maturitätsprüfung 2007 = w_{tG} $= \frac{1}{n_G} \sum_{j=1}^{n_G} \sum_{a_{ij} \in M_t} \frac{a_{ij}}{z_j}$

Die Resultate der Analyse finden sich in Tabelle V.29.

Tab. V.29: Durchschnittliches Gewicht des Teilbereichs *t* auf Note oder Gesamtpunktzahl

	Cytologie/ Anatomie/ Stoffwechsel	Informations-verarbeitung, Verhalten zusammen mit Immunbiologie	Genetik/ Entwick-lungsbiologie	Ökologie	Systematik/ Evolution	divers/ vermischt	Total
Mittleres relatives Kapitelvolumen in drei deutschsprachigen Standardlehrwerken der Biologie auf der Gymnasialstufe (Stufe SII)	23.1%	24.2%	19.6%	17.3%	15.8%	0%	100.0%
I durchschnittlicher Anteil der einzelnen Teilbereiche an der Gesamtpunktzahl in Maturitätsprüfungen im Fach Biologie 2007 w_{tG}	28.5%	21.9%	26.2%	10.2%	6.4%	6.9%	100.1%
II korrigierte Prozentsätze aus I nach proportionaler Aufteilung des Anteils „divers/vermischt“	30.6%	23.5%	28.1%	10.9%	6.9%	0%	100.0%
Anteile Maturitätsprüfungen (korrigierte Prozentsätze) vs. Standardlehrwerke (hier als 100%-Wert genommen)	132%	97%	143%	63%	44%		

Im Vergleich mit der Gewichtung in den Lehrwerken kann eine Probandin oder ein Proband an einer durchschnittlichen Maturitätsprüfung im Fach Biologie die Gesamtpunktzahl besonders effizient mit Fähigkeiten in den Teilbereichen Cytologie/Anatomie/Stoffwechsel sowie auch Genetik/Entwicklungsbiologie beeinflussen. Hingegen sind die Teilbereiche Ökologie und Systematik/Evolution in den Standardlehrbüchern deutlich gewichtiger repräsentiert als an einer durchschnittlichen Maturitätsprüfung im Fach Biologie. Dies könnte damit zusammenhängen, dass die Behandlung der Evolutionstheorie am Gymnasium häufig hintangestellt wird. Die Vernachlässigung des Themas könnte ein Grund dafür gewesen sein, dass die Items zum Thema Evolutionstheorie auch von guten Maturandinnen und Maturanden relativ schlecht gelöst wurden, so dass viele Items aus diesem Themenbereich die nötige minimale Diskriminationsstärke für die Aufnahme in die „Itemselektion Biologie EVAMAR II" (endgültige Auswahl für die Hauptauswertungen; siehe Abschnitt 2.4. in Kapitel IV) nicht erreichten.

3.3.3 Kognitive Anforderungen

Werden die Maturandinnen und Maturanden auf die kognitiven Anforderungen der Prüfungsaufgaben an der Universität vorbereitet? Während diese Frage für die in EVAMAR II untersuchten Fächer Mathematik und Erstsprache schwierig zu untersuchen ist – beide Fächer bereiten auf Anforderungen in vielen Studienfächern vor und es müssten aus den entsprechenden Prüfungen die Mathematik- und Sprachanteile herausgefiltert werden –, lässt sich dies bei Biologie besser feststellen. Dieses Maturafach ist im Vergleich zu den anderen beiden am ehesten als direkte Vorbereitung auf Fachbiologie an der Universität zu sehen und kann deshalb direkt mit universitären Biologie-Prüfungen verglichen werden. Die Frage nach einer adäquaten Vorbereitung lässt sich beantworten, indem man das kognitive Aktivierungspotenzial von typischen Prüfungsaufgaben beider Ausbildungsstufen miteinander vergleicht. Eine häufig genannte Hypothese lautet dahin gehend, dass Biologieaufgaben in den schweizerischen Maturitätsprüfungen 2007 ein höheres kognitives Aktivierungspotenzial als die Aufgaben aus Hochschulprüfungen haben. Begründet wird diese Hypothese in der Regel aus der Überlegung heraus, dass aufgrund der hohen Studierendenzahlen pro Veranstaltung und der hohen Arbeitsbelastung der Dozierenden der Korrekturaufwand systembedingt stark begrenzt werden muss und ein grosser Anteil von Aufgaben im effizient korrigierbaren Multiple-Choice-Format formuliert sein dürfte. Gegen Fragestellungen im Essayformat mit hohem Aktivierungspotenzial spricht, dass der Korrekturschlüssel mit jeder Essayfrage, die anspruchsvolle Transfer- und Metakognitionsleistungen verlangt, weniger objektivierbar wird. Dies bringt einerseits den Nachteil der leichteren Anfechtbarkeit der Korrektur mit sich, was bei der enormen Wichtigkeit der Prüfungsresultate für den einzelnen Prüfling ein gewisses Risiko für die Prüfenden darstellt. Andererseits lassen sich wenig objektivierbare Korrekturschlüssel schlecht mit anderen Personen teilen. Es ist für den Dozierenden schwierig, Teile der Korrekturarbeit zu delegieren, wenn nur ein schwach objektivierbarer Korrekturschlüssel möglich ist.

Für die Analyse wurde zunächst das durchschnittliche notenrelevante Gewicht der Aufgaben mit einem bestimmten kognitiven Aktivierungspotenzial an einer Maturitätsprüfung ermittelt. Die Gewichte für jeden Aufgabentyp wurden durch folgende Berechnung geschätzt:

- geg. Maturitätsprüfung j
- geg. Prüfungsteilfrage Q_{ij} = Prüfungsfrage i innerhalb Maturitätsprüfung j
- geg. n_G = Gesamtzahl der analysierten Maturitätsprüfungen = 65
- geg. z_j = maximal erreichbare Punktzahl für das Beantworten der ganzen Maturitätsprüfung j
- geg. a_{ij} = maximal erreichbare Punktzahl für das Beantworten der Prüfungsfrage Q_{ij}
- M_p = Menge aller a_{ij}, so dass die entsprechenden Prüfungsfragen Q_{ij} das Aktivierungspotenzial p haben (gemäss der Inhaltsanalyse)

- ges. durchschnittliches notenrelevantes Gewicht der Aufgaben mit kognitivem Aktivierungspotenzial p an einer Maturitätsprüfung = $w_{pG} = \frac{1}{n_G} \sum_{j=1}^{n_G} \sum_{a_{ij} \in M_p} \frac{a_{ij}}{z_j}$

Die Ergebnisse für w_{pG}-Werte finden sich in Tabelle V.30.

Tab. V.30: Anteil der Gesamtpunktzahl, der durch das Lösen von Aufgaben einer gewissen Kategorie in einer durchschnittlichen Biologie-Maturitätsprüfung erreicht werden kann.

Hauptkategorie und Unterkategorien	In EVAMAR II für das Fach Biologie adaptierte Definitionen *[angelehnt an Anderson et al. (2001), p. 63–92]*	Anteil der Gesamtpunktzahl erreichbar durch Aufgaben dieser Kategorie w_{pG} in Prozent
1	Erinnern relevantes Wissen aus dem Langzeitgedächtnis abrufen	
1.1	Wissen, das mit dem vorgegebenen Material konsistent ist, im Langzeitgedächtnis auffinden	1.6%
1.2	relevantes Wissen aus dem Langzeitgedächtnis abrufen	47.0%
2	Verstehen aus Lehrinhalten (mündlich, schriftlich und grafisch) Bedeutung gewinnen	
2.1	vorgegebenes Material von einer Darstellungsform in eine andere bringen	1.4%
2.2	ein bestimmtes Beispiel oder eine bestimmte Abbildung finden, welche(s) ein Konzept oder ein Prinzip veranschaulicht	0.6%
2.3	entscheiden, welche Instanzen zu welchen Kategorien (Konzepten oder Prinzipien) gehören	1.3%
2.4	zu vorgegebenem Material ein Hauptthema finden oder die wichtigsten Punkte zusammenfassen (mit eigener Integrationsleistung) aus dem vorgelegten Material die Quintessenz, die Pointe, das „Worum es geht" herausfiltern und zusammenfassen (mit eigener Integrationsleistung)	0.5%
2.6	Übereinstimmungen und Unterschiede zwischen Ideen, Objekten u.Ä. erkennen Ideen, Objekte u.Ä. innerhalb derselben Hierarchie-Ebene einander zuordnen (Mapping)	5.9%
2.7	ein zuvor erlerntes Ursache-Wirkungs-Modell auf bisher unbekannte Tatsachen anwenden Anwenden heisst: geg. Wirkung(en): das Modell zum Schliessen auf Ursachen verwenden geg. Ursache(n): das Modell zum Schliessen auf Wirkungen verwenden	20.6%
3	Anwenden in einer gegebenen Situation eine Prozedur ausführen oder einsetzen	
3.2	eine Prozedur (mechanisch einsetzbare und einübbare Schritte) anpassen, damit sie auf ein nicht bereits vertrautes Problem angewendet werden kann	11.1%
4	Analysieren vorgegebenes Material in seine Bestandteile zerlegen und bestimmen, wie die Teile untereinander in Beziehung stehen und wie die Teile mit der Gesamtstruktur oder einem Gesamtzweck zusammenhängen	
4.1	relevante (bzw. wichtige) Teile und irrelevante (bzw. unwichtige) Teile innerhalb des vorgegebenen Materials unterscheiden (ohne eine eigene Integrationsleistung)	0.9%
4.2	bestimmen, wie Elemente innerhalb einer Struktur zusammenpassen oder	1.5%

Hauptkategorie und Unterkategorien	In EVAMAR II für das Fach Biologie adaptierte Definitionen *[angelehnt an Anderson et al. (2001), p. 63–92]*	Anteil der Gesamtpunktzahl erreichbar durch Aufgaben dieser Kategorie w_{pG} in Prozent
	zusammenspielen vorgelegtes Material strukturieren (Zwischentitel setzen, Tabelle konstruieren, Box-Pfeil-Diagramm oder Mindmap erstellen usw.), also das Material auch in eine andere Form bringen.	
4.3	Blickwinkel, Wahrnehmungsverzerrungen, Wertvorstellungen oder Absichten des Autors des vorgegebenen Materials erkennen	0.2%
5	Bewerten auf Kriterien und Standards abgestützt ein Urteil abgeben	
5.1	Inkonsistenzen oder Fehler innerhalb eines Prozesses oder eines Produktes erkennen; entscheiden, ob ein Prozess oder Produkt (in sich selber) konsistent ist; die Effektivität einer Prozedur (so wie sie angewendet wird) bestimmen	2.9%
5.2	erkennen, wo ein Prozess oder ein Produkt inkonsistent ist gemessen an externen Kriterien; bestimmen, ob ein Produkt externe Konsistenz hat; erkennen, ob eine Prozedur bei einem gegebenen Problem angemessen ist	3.1%
6	Erschaffen Elemente zusammensetzen, um ein kohärentes funktionales Ganzes zu formen; Elemente auf neuartige Weise strukturieren	
6.1	sich Hypothesen ausdenken	1.3%
6.2	eine Vorgehensweise entwickeln, um eine Aufgabe zu lösen	0.2%
6.3	ein Produkt neu erfinden	0.0%

Drei Eigenschaften der Maturitätsprüfungen 2007 im Fach Biologie fallen auf:

a) Ungefähr die Hälfte der Gesamtpunktzahl (48.6%) erreicht in einer durchschnittlichen Maturitätsprüfung im Fach Biologie, wer ausschliesslich reine Wissensaufgaben lösen kann (Kategorien 1.1 und 1.2), welche über den Gedächtniszugriff hinaus keine weitergehenden kognitiven Prozesse erfordern.

b) Maturandinnen und Maturanden, welche die folgenden vier dominanten kognitiven Prozesse gut beherrschen, erreichen bereits 84.4% der Gesamtpunktzahl:

- relevantes Wissen aus dem Langzeitgedächtnis abrufen (1.2),
- Übereinstimmungen und Unterschiede zwischen Ideen, Objekten u.Ä. erkennen und Ideen, Objekte u.Ä. innerhalb derselben Hierarchie-Ebene einander zuordnen (Mapping) (2.6),
- ein zuvor erlerntes Ursache-Wirkungs-Modell auf bisher unbekannte Tatsachen anwenden (2.7),
- eine Prozedur (mechanisch einsetzbare und einübbare Schritte) anpassen, damit sie auf ein nicht bereits vertrautes Problem angewendet werden kann (3.1/3.2).

Die Beherrschung aller übrigen 17 kognitiven Prozesse erbringt zusammen genommen 15.6% der Gesamtpunktzahl. Die Verteilung ist also sehr ungleichmässig.

c) Im Sinne einer Bestimmung des Punktebeitrags von Aufgaben, die potenziell Prozesse des Transfers und der Metakognition aktivieren (Georghiades, 2000), können die durchschnittlichen Punktebeiträge der Hauptkategorien 3 bis 6 addiert werden (Anderson et al., 2001, S. 63). Auch die der Unterkategorie 2.7 zugeteilten Aufgaben aktivieren gemäss der im Rahmen von EVAMAR II gewählten Definition eindeutig Transferleistungen. Ihr durchschnittlicher Punktebeitrag muss zu dem gesuchten Punktebeitrag addiert werden. Der kognitive Prozess des Vergleichens (Unterkategorie 2.6) wird hier nicht zu den Transferleistungen gezählt. Nach dieser Aufteilung liegt der Anteil der Gesamtpunktzahl, welcher durch das Lösen von transfer- und metakognitionslastigen Aufgaben erreicht werden kann, bei 42%.

Aus Abbildung V.1 ist ersichtlich, dass die Anteile an der Gesamtpunktzahl sehr stark gehäuft verteilt sind. Mit dem Lösen der vier gewichtigsten Aufgabenkategorien kann bereits 84.4% der Gesamtpunktzahl einer durchschnittlichen Maturitätsprüfung in Biologie erreicht werden.

Abb. V.1: Anteile der verschiedenen Kategorien an der Gesamtpunktzahl N = 65
(54 Maturaprüfungen, 9 Vor-Maturaprüfungen, 2 Ergänzungswahlfachprüfungen; D 58, F 4, I 3)

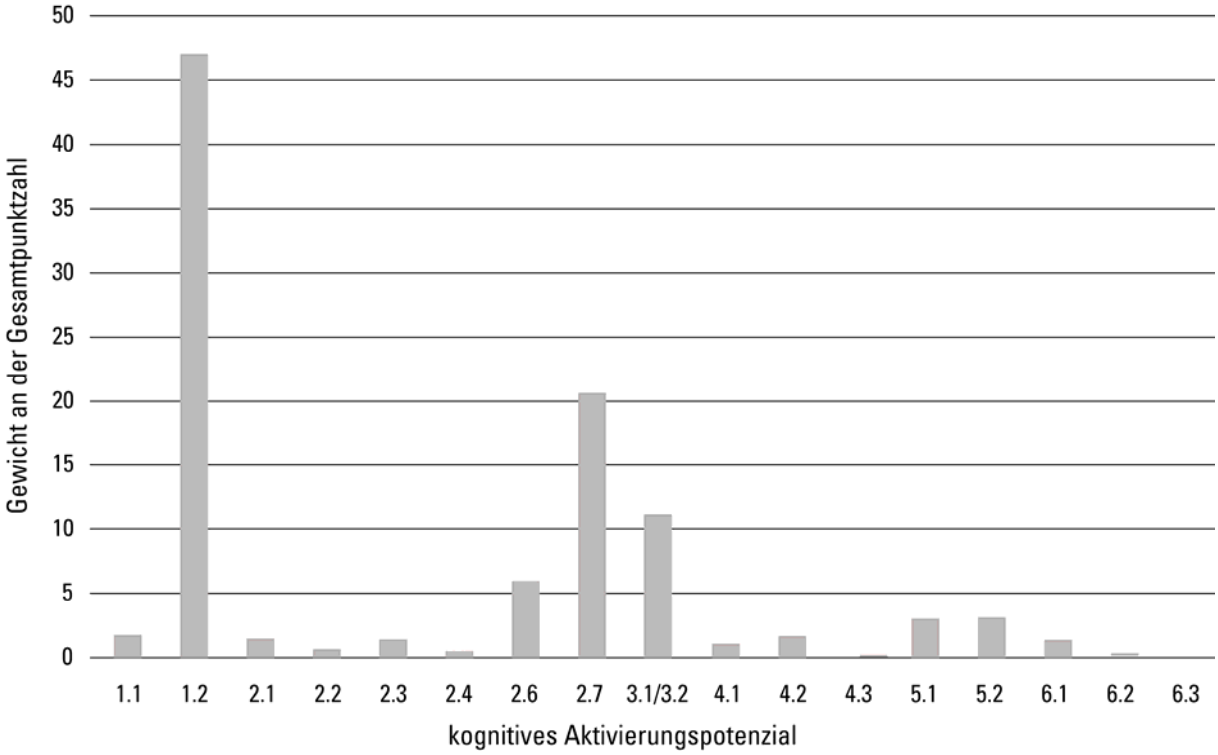
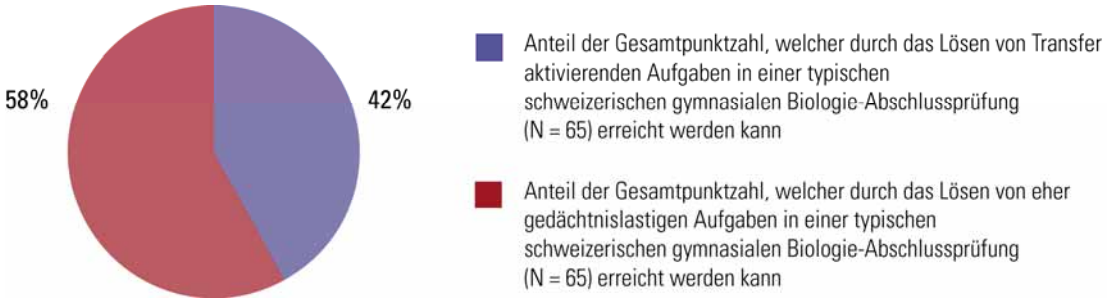


Abbildung V.2 zeigt im Weiteren, dass allein durch das Lösen aller eher gedächtnislastigen Aufgaben an einer durchschnittlichen Maturitätsprüfung in Biologie erst 58% der Gesamtpunktzahl erreicht werden kann. Die restlichen Punkte müssen mit Transferleistungen geholt werden.

Abb. V.2: Anteile von Erinnerungs- und Transferaufgaben (Biologie, Gymnasium)



Zur Untersuchung der Hochschulprüfungen in Biologie wurde grundsätzlich dieselbe Berechnungsmethode wie bei den Maturitätsprüfungen angewendet. Die Gesamtzahl der analysierten Prüfungen liegt aber lediglich bei n = 26, da nicht sämtliche Prüfungen zur Verfügung gestellt wurden. Es lagen 19 Prüfungen aus dem ersten Semester des Biologiestudiums (Biologische Richtung) an der ETHZ und 7 Prüfungen aus den ersten beiden Semestern an der Universität Zürich vor. Da die Stichprobengrösse für die Universität Zürich zu klein und die Stichproben beider Hochschulen nicht randomisiert waren, wurden die beiden Hochschulen nicht verglichen, sondern gemeinsam betrachtet, wobei jede Hochschule mit demselben Gewicht von 1.0 in die Gesamtbetrachtung einfloss

(siehe die Spalte rechts aussen in Tab. V.31). An der ETHZ haben die Veranstaltungen unterschiedliches Notengewicht. Diesem Umstand wurde hier Rechnung getragen. Da aber nicht alle Prüfungen aus sämtlichen Veranstaltungen des 1. Semesters vorlagen, wurde das Gesamtgewicht der ETHZ proportional zu den Notengewichten aus der Prüfungsverordnung Bachelorstudium 2005/06 auf die drei analysierten Veranstaltungen aufgeteilt. Veranstaltungen, deren Aufgabentypen durch eine relativ grosse Zahl von vorliegenden Prüfungen analysiert werden konnten, flossen selbstverständlich nicht gewichtiger in die Gesamtbeurteilung ein als Veranstaltungen, die durch eine geringere Anzahl Prüfungen dokumentiert sind. Dies ist an der Drittelung des Gewichts der Veranstaltung „GL der Biologie IA: Allgemeine Biologie" an der ETHZ zu sehen. An der Universität Zürich muss jede Modulprüfung mit einer genügenden Note bestanden werden. Die Modulprüfungen haben somit dasselbe Gewicht. Da Prüfungen aus fünf Modulen zur Analyse vorlagen, wurde das Gesamtgewicht der Universität Zürich in fünf Teile geteilt: Jedes Modul erhielt das Gewicht von 0.2. Die Berechnung erfolgte im Einzelnen wie folgt:

- geg. Hochschulprüfung j
- geg. Prüfungsteilfrage Q_{ij} = Prüfungsfrage i innerhalb Hochschulprüfung j
- geg. n_H = Gesamtzahl der analysierten Hochschulprüfungen = 26
- geg. z_j = maximal erreichbare Punktzahl für das Beantworten der ganzen Hochschulprüfung j
- geg. a_{ij} = maximal erreichbare Punktzahl für das Beantworten der Prüfungsfrage Q_{ij}
- M_p = Menge aller a_{ij}, so dass die entsprechenden Prüfungsfragen Q_{ij} das Aktivierungspotenzial p haben (gemäss der Inhaltsanalyse)
- e_{ij} = Anzahl analysierte Prüfungen derjenigen Veranstaltung, zu der die Aufgabe Q_{ij} gehört
- d_{ij} = Anzahl Dozierende, die diejenige Veranstaltung bestreiten, zu der die Aufgabe Q_{ij} gehört
- w_{reg} = relatives Gewicht der Veranstaltung gemäss Prüfungsreglement, wenn das Gesamtgewicht der analysierten Prüfungen pro Hochschule 1.0 beträgt
- ges. durchschnittliches notenrelevantes Gewicht der Aufgaben mit kognitivem Aktivierungspotenzial p an einer Hochschulprüfung = w_{pH} = $\sum_{j=1}^{n_H} \sum_{a_{ij} \in M_p} \frac{a_{ij}}{z_j} \cdot \frac{1}{e_{ij}} \cdot \frac{1}{d_{ij}} \cdot w_{reg}$

Eine Übersicht über die analysierten Hochschulprüfungen findet sich im technischen Bericht. Die Resultate der Analyse sind in Tabelle V.31 zusammengestellt. Sie zeigt den Anteil der Gesamtpunktzahl, der durch das Lösen von Aufgaben einer gewissen Kategorie in einer durchschnittlichen Prüfung an der Hochschule in den beiden ersten Semestern erreicht werden kann.

Tab. V.31: Vereilung der Gesamtpunktzahl von Hochschulprüfungen in Biologie auf die Kategorien des kognitiven Aktivierungspotenzials

Hauptkategorie und Unterkategorien	In EVAMAR II für das Fach Biologie adaptierte Definitionen	Anteil der Gesamtpunktzahl erreichbar durch Aufgaben dieser Kategorie
	[angelehnt an Anderson et al. (2001), p. 63–92]	w_{pH} in Prozent
1	Erinnern relevantes Wissen aus dem Langzeitgedächtnis abrufen	
1.1	Wissen, das mit dem vorgegebenen Material konsistent ist, im Langzeitgedächtnis auffinden	8.3%
1.2	relevantes Wissen aus dem Langzeitgedächtnis abrufen	39.2%
2	Verstehen aus Lehrinhalten (mündlich, schriftlich und grafisch) Bedeutung gewinnen	

Hauptkategorie und Unterkategorien	In EVAMAR II für das Fach Biologie adaptierte Definitionen	Anteil der Gesamtpunktzahl erreichbar durch Aufgaben dieser Kategorie
	[angelehnt an Anderson et al. (2001), p. 63–92]	w_{pH} in Prozent
2.1	vorgegebenes Material von einer Darstellungsform in eine andere bringen	0.0%
2.2	ein bestimmtes Beispiel oder eine bestimmte Abbildung finden, welche(s) ein Konzept oder ein Prinzip veranschaulicht	0.0%
2.3	entscheiden, welche Instanzen zu welchen Kategorien (Konzepten oder Prinzipien) gehören	8.1%
2.4	zu vorgegebenem Material ein Hauptthema finden oder die wichtigsten Punkte zusammenfassen (mit eigener Integrationsleistung) aus dem vorgelegten Material die Quintessenz, die Pointe, das „Worum es geht" herausfiltern und zusammenfassen (mit eigener Integrationsleistung)	0.0%
2.6	Übereinstimmungen und Unterschiede zwischen Ideen, Objekten u.Ä. erkennen Ideen, Objekte u.Ä. innerhalb derselben Hierarchie-Ebene einander zuordnen (Mapping)	5.8%
2.7	ein zuvor erlerntes Ursache-Wirkungs-Modell auf bisher unbekannte Tatsachen anwenden Anwenden heisst: geg. Wirkung(en): das Modell zum Schliessen auf Ursachen verwenden geg. Ursache(n): das Modell zum Schliessen auf Wirkungen verwenden	6.8%
3	Anwenden in einer gegebenen Situation eine Prozedur ausführen oder einsetzen	
3.2	eine Prozedur (mechanisch einsetzbare und einübbare Schritte) anpassen, damit sie auf ein nicht bereits vertrautes Problem angewendet werden kann	6.2%
4	Analysieren vorgegebenes Material in seine Bestandteile zerlegen und bestimmen, wie die Teile untereinander in Beziehung stehen und wie die Teile mit der Gesamtstruktur oder einem Gesamtzweck zusammenhängen	
4.1	relevante (bzw. wichtige) Teile und irrelevante (bzw. unwichtige) Teile innerhalb des vorgegebenen Materials unterscheiden (ohne eigene Integrationsleistung)	1.0%
4.2	bestimmen, wie Elemente innerhalb einer Struktur zusammenpassen oder zusammenspielen vorgelegtes Material strukturieren (Zwischentitel setzen, Tabelle konstruieren, Box-Pfeil-Diagramm oder Mindmap erstellen usw.), also das Material auch in eine andere Form bringen	2.9%
4.3	Blickwinkel, Wahrnehmungsverzerrungen, Wertvorstellungen oder Absichten des Autors des vorgegebenen Materials erkennen	0.0%
5	Bewerten auf Kriterien und Standards abgestützt ein Urteil abgeben	
5.1	Inkonsistenzen oder Fehler innerhalb eines Prozesses oder eines Produktes erkennen; entscheiden, ob ein Prozess oder Produkt (in sich selber) konsistent ist; die Effektivität einer Prozedur (so wie sie angewendet wird) bestimmen	22.5%
5.2	erkennen, wo ein Prozess oder ein Produkt inkonsistent ist gemessen an externen Kriterien; bestimmen, ob ein Produkt externe Konsistenz hat; erkennen, ob eine Prozedur bei einem gegebenen Problem angemessen ist	0.0%
6	Erschaffen Elemente zusammensetzen, um ein kohärentes funktionales Ganzes zu formen; Elemente auf neuartige Weise strukturieren	

Hauptkategorie und Unterkategorien	In EVAMAR II für das Fach Biologie adaptierte Definitionen	Anteil der Gesamtpunktzahl erreichbar durch Aufgaben dieser Kategorie
	[angelehnt an Anderson et al. (2001), p. 63–92]	w_{pH} in Prozent
6.1	sich Hypothesen ausdenken	0.1%
6.2	eine Vorgehensweise entwickeln, um eine Aufgabe zu lösen	0.0%
6.3	ein Produkt neu erfinden	0.0%

Die drei wichtigsten Ergebnisse sind die Folgenden:

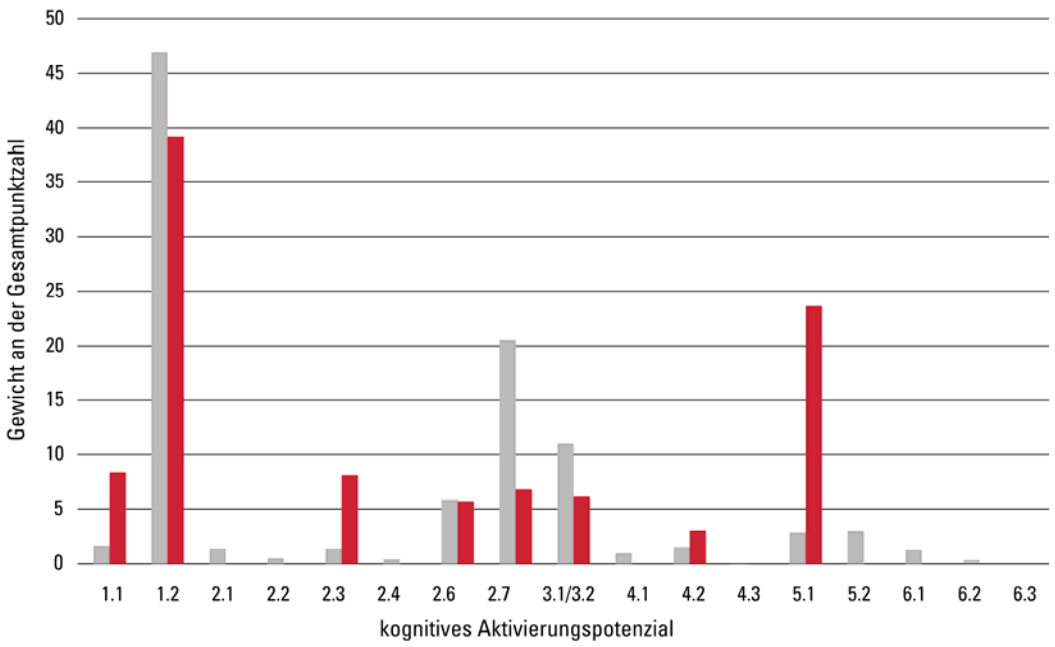
a) Wiederum ungefähr die Hälfte der Gesamtpunktzahl, nämlich 47.6%, erreicht in einer durchschnittlichen Hochschulprüfung im Fach Biologie, wer ausschliesslich reine Wissensaufgaben lösen kann (Kategorien 1.1 und 1.2), also solche Aufgaben, die über den Gedächtniszugriff hinaus keine weitergehenden kognitiven Prozesse erfordern.

b) Die Verteilung ist wie bei den Maturitätsprüfungen sehr ungleichmässig. Studentinnen und Studenten, welche die folgenden vier dominanten kognitiven Prozesse gut beherrschen, erreichen bereits 78% der Gesamtpunktzahl:

- Wissen, das mit dem vorgegebenen Material konsistent ist, im Langzeitgedächtnis auffinden (1.1),
- relevantes Wissen aus dem Langzeitgedächtnis abrufen (1.2),
- entscheiden, welche Instanzen zu welchen Kategorien (Konzepten oder Prinzipien) gehören (2.3),
- Inkonsistenzen oder Fehler innerhalb eines Prozesses oder eines Produktes erkennen; entscheiden, ob ein Prozess oder Produkt (in sich selber) konsistent ist; die Effektivität einer Prozedur (so wie sie angewendet wird) bestimmen (5.1).

c) Der Anteil der Gesamtpunktzahl, welcher durch das Lösen von transfer- und metakognitionslastigen Aufgaben erreicht werden kann, liegt bei 38.5%, also nur geringfügig unter dem entsprechenden Wert von 42.2% bei den Maturitätsprüfungen (siehe Abb. V.3).

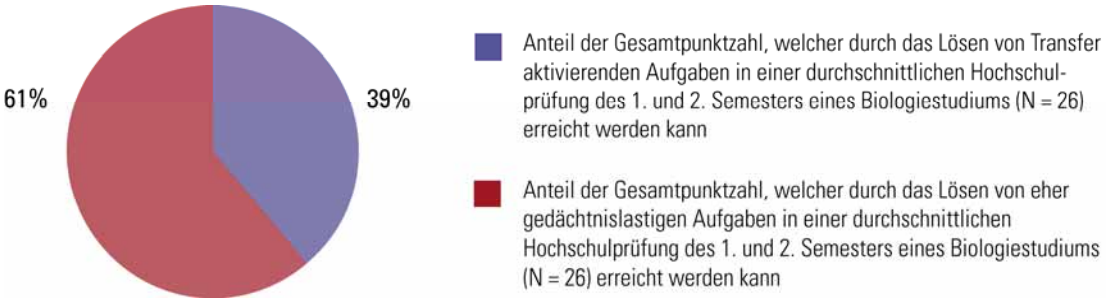
Abb. V.3: Vergleich der kognitiven Anforderungen von Maturitätsprüfungen und universitären Prüfungen in Biologie



Beim Vergleich der kognitiven Anforderungspotenziale zwischen Maturitätsprüfungen (schräg schraffiert) und Hochschulprüfungen (kariert) fällt weiter auf, dass an den Hochschulen Aufgaben der Kategorie 5.1 (Inkonsistenzen und Fehler erkennen) sehr viel häufiger sind als an den Gymnasien. Dafür handelt es sich bei Aufgaben der Kategorie 2.7 (ein zuvor erlerntes Ursache-Wirkungs-Modell auf bisher unbekannte Tatsachen anwenden) offenbar um eine „Spezialität" der Maturitätsprüfungen.

Allein durch das Lösen aller eher gedächtnislastigen Aufgaben kann an einer durchschnittlichen Hochschulprüfung im Studienfach Biologie erst 61.5% der Gesamtpunktzahl erreicht werden. Die restlichen Punkte müssen mit Transferleistungen geholt werden (siehe Abb. V.4).

Abb. V.4: Anteile von Erinnerungs- und Transferaufgaben (Biologie, Hochschule)



Aus dem Vergleich von Maturitätsprüfungen und universitären Prüfungen lässt sich folgendes Gesamtfazit ziehen: Hochschulprüfungen des Studienfachs Biologie enthalten zwar viele Multiple-Choice-Aufgaben, aber drei Viertel von ihnen können nicht mit einer simplen Wiedererkennungsleistung gelöst werden. Die Optionen dieser anspruchsvollen Multiple-Choice-Aufgaben bestehen aus ganzen Sätzen, deren einzelne Richtigkeit oder Fehlerhaftigkeit erst aus dem Gesamtverständnis des ganzen Satzes heraus entschieden werden kann. Damit überprüfen diese Fragen die Tiefe des Verständnisses und dürfen zu den Aufgaben mit einem hohen kognitiven Aktivierungspotenzial gezählt werden. Annähernd 40% der Gesamtpunktzahl an einer Hochschulprüfung im Studienfach Biologie müssen also durch das Lösen von Aufgaben erarbeitet werden, die nicht durch eine oberflächliche Gedächtnisleistung gemeistert werden können. Dieser Prozentsatz ist annähernd so hoch wie derjenige bei den schweizerischen Maturitätsprüfungen 2007 (42%). Dies erstaunt angesichts der oben genannten Gründe für die Formulierung der Hypothese sehr. Offenbar haben viele Dozierende mit der Formulierung anspruchsvoller Multiple-Choice-Aufgaben einen Weg gefunden, Korrekturaufwand und -objektivierbarkeit sowie das Prüfen der Verständnistiefe unter einen Hut zu bringen. Tatsächlich revidiert die Forschung im Bereich der Testentwicklung immer mehr das Vorurteil, Multiple-Choice-Aufgaben eigneten sich nur zum Abfragen von Faktenwissen (als Review z. B. Haladyna, Downing & Rodriguez, 2002).

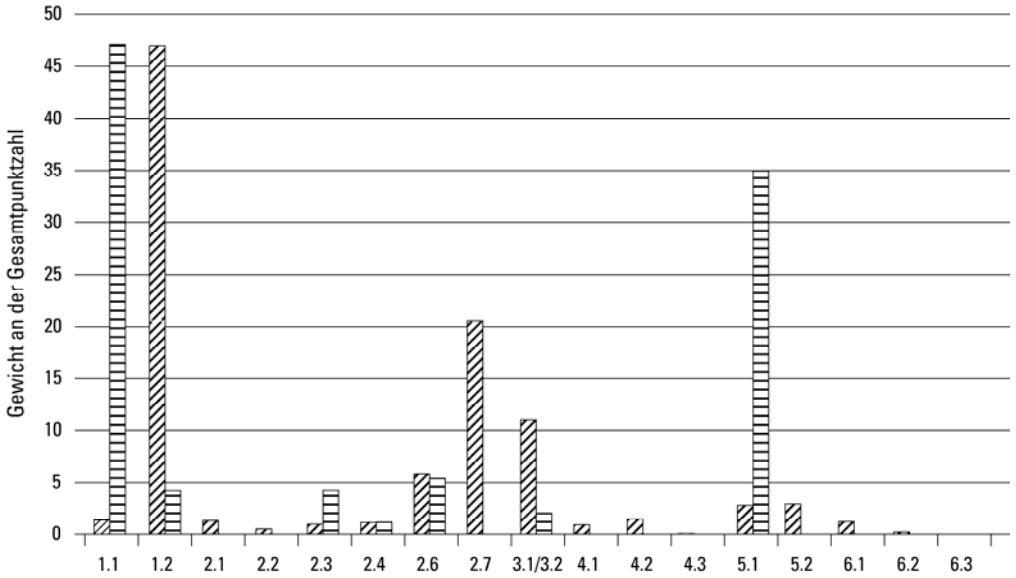
3.3.4 Vergleich mit den Aufgaben in Teilprojekt C

Die Items des Biologie-Tests in Teilprojekt C sollten vor allem messen, in welchem Ausmass eine Probandin bzw. ein Proband über jenes biologische Fachwissen und -können verfügt, um Inhalten über Biologie in Lehrtexten aus dem ersten Semester ohne Hilfsmittel folgen zu können. Die Formulierung der meisten Items im Multiple-Choice-Format zielte auf eine Minimierung des Korrekturaufwands. Zu Vergleichszwecken wurde das kognitive Aktivierungspotenzial der in die Hauptauswertungen einbezogene Items („Item-Selektion Biologie EVAMAR II") auf dieselbe Weise wie die Aufgaben der Maturitätsprüfungen und die Aufgaben aus den Hochschulprüfungen eingeschätzt (siehe Tab. V.32).

Tab. V.32: Anteil der Items innerhalb der Item-Selektion Biologie EVAMAR II mit einem bestimmten kognitiven Aktivierungspotenzial

Kategorie	Format des Items	Anzahl Items	Anteil w_{pE}
1.1	*Multiple-Choice* Mit stichwortartigen Optionen; Optionen enthalten eine Lehrbuch-Definition, d. h. direktes Wiedererkennen der korrekten Option ist möglich. Der Stamm der Aufgabe ist eine Frage. Er macht keine Aussage.	43	47.3%
1.2	*offen* Fachbegriff aus Langzeitgedächtnis erinnern	4	4.4%
2.3	*Multiple-Choice* Nur eine Option ist die korrekte Instanz oder das korrekte Konzept. Gegebene Instanzen korrekt gegebenen Konzepten zuordnen.	4	4.4%
2.4	*offen* Gegeben ist ein Diagramm, gesucht ist der Diagrammtyp.	1	1.1%
2.6	*Multiple-Choice* Die Optionen vergleichen zwei Konzepte, nur ein Vergleich ist korrekt.	5	5.5%
3.1	*Multiple-Choice* Anwenden eines Schemas auf neue Abbildungen, z. B. Lesen eines Stammbaumes bei korrekter Anwendung kann die korrekte Option gewählt werden.	2	2.2%
5.1	*Multiple-Choice* Der Stamm ist eine Aussage aus dem Studienmaterial plus eine Frage. Es stehen drei mit der Aussage inkonsistente Optionen und eine mit der Aussage konsistente Option zur Auswahl. Das Überprüfen der Konsistenz erfordert biologisches Verständnis: Was bedeuten die Aussagen?	32	35.2%
Total der Items und Prozentsumme		91	100.1%

Abb. V.5: Verteilung der kognitiven Aktivierungspotenziale der Aufgaben aus schweizerischen Maturitätsprüfungen 2007 (schräg schraffiert) mit denjenigen Test-Items aus der Item-Selektion Biologie EVAMAR II (waagrecht schraffiert)



In Abbildung V.5 werden Maturaaufgaben mit den Aufgaben des Biologietests verglichen. Es ist sofort zu erkennen, dass die Items des Biologietests von EVAMAR II reine Gedächtnisleistung nicht offen abfragen, sondern mithilfe eines Multiple-Choice-Items, das stichwortartige Optionen enthält und die Möglichkeit eröffnet, die richtige Lösung nur durch Wiedererkennen zu finden. In den Maturitätsprüfungen muss die Gedächtnisleistung wesentlich häufiger aktiv erfolgen. Hieran ist gut zu erkennen, dass bei den rund 3'800 Probanden der Hauptuntersuchung die Ökonomie der Korrektur im Vordergrund stand und stehen musste. Man sieht zudem, dass Aufgaben, welche die Tiefe des Verständnisses prüfen, in den Maturitätsprüfungen häufig in Form einer Aufgabe mit dem kognitiven Aktivierungspotenzial 2.7 formuliert werden. Da dieses Aufgabenformat meist zu mehrsätzigen und aufwändig zu korrigierenden Antwort-Essays führt, wurde bei der Konstruktion der Test-Items derselbe Weg beschritten, den die Hochschulen offenbar auch wählen. Viele der die Verständnistiefe testenden Items innerhalb der Item-Selektion Biologie EVAMAR II sind Multiple-Choice-Fragen, die das Potenzial haben, den anspruchsvollen kognitiven Prozess des Fehlererkennens zu aktivieren.

3.3.5 Vergleich mit den Ergebnissen aus Teilprojekt A4

In Tabelle V.33 sind die Äusserungen von Dozentinnen und Dozenten der Studienfächer Biologie und Medizin bezüglich festgestellter Wissens- und Könnenslücken der Studienanfängerinnen und -anfänger in Biologie dargestellt. Nicht weniger als 6 von 21 Biologie- und Medizindozentinnen und -dozenten, die den Fragebogen zurückgesandt haben, drücken durch ihre Antworten aus, dass sie sich bessere Kenntnisse in den Teilbereichen Ökologie und Systematik/Evolution wünschen. Nur 2 dieser 21 Dozierenden wünschen sich mehr Kenntnisse in den anderen Teilgebieten. Dozentinnen und Dozenten der Biologie und Medizin erkennen also bei denjenigen Teilbereichen, die bei deutsch- und französischsprachigen Maturitätsprüfungen eher ein geringes Gewicht haben, die grössten Defizite. Ob die inhaltliche Verteilung der Prüfungsfragen etwas über die inhaltliche Verteilung der Unterrichtszeit oder die Schwierigkeit der Teilgebiete aussagt, ist damit nicht untersucht.

Tab. V.33: Festgestellte Lücken in Biologie und gewünschtes zusätzliches Eingangswissen und -können (Ökologie und Systematik/Evolution kursiv)

Nr. Dozent	Frage 3: Aktuell festgestellte Wissens- und Könnenslücken	Zusatzfrage 6: Wünschbares zusätzliches Eingangswissen und -können
23	keine Ahnung, was Fotosynthese ist sehr spärliches Wissen, was bei Mitose und Meiose passiert *Grundkenntnisse in Ökologie mangelhaft und oftmals gar nicht vorhanden*	*mehr ökologische Fragen und Probleme behandeln*
39	*generell scheint das evolutionsbiologische Verständnis/Wissen am marginalsten zu sein*	
55	chemisches Verständnis der Biologie	
78	*eigene Erfahrungen aus dem Freiland (dank Exkursionen, Semester-/ Matura-Arbeiten im Feld u.Ä.)* *Formenkenntnisse* *Kenntnis der aktuellen Biodiversitätskrise und der wichtigsten Gründe für den Artenrückgang in Mitteleuropa* *teilweise auch* *Verständnis für Stammbäume*	*eigene Erfahrungen aus dem Freiland (dank Exkursionen, Semester-/ Matura-Arbeiten im Feld u.Ä.)* *Formenkenntnisse* *Kenntnis der aktuellen Biodiversitätskrise und der wichtigsten Gründe für den Artenrückgang in Mitteleuropa*
106		*Ökologie: Wissen um die ökologischen Probleme der Schweiz und der Welt*
51		In der Biologie sind nicht so sehr die Einzelfakten moderner Zell- und Molekularbiologie zu vermitteln, als vielmehr *Stoffkreisläufe, ökologische Zusammenhänge, Artenvielfalt und Evolution.*

3.3.6 Gesamteinschätzung der Schweizer Maturaprüfungen in Biologie

In der Literatur findet sich gemäss unserem Wissen nur gerade eine Studie, welche die Bedeutung der einzelnen Aufgabentypen ebenfalls entsprechend dem Einfluss auf die Gesamtpunktzahl gewichtet und sich auf Prüfungen im Fach Biologie bezieht. Die Studie findet Anteile von „higher-order questions“ (Hauptkategorien 3 bis 6 bei Bloom) an der Gesamtpunktzahl zwischen 21 und 52% (Zheng et al., 2008). Allerdings sind die untersuchten Prüfungen nicht für den Gebrauch im Klassenverband ausgelegt, sondern es handelt sich um bekannte, gross angelegte US-amerikanische Assessments (der Medical College Admission Test MCAT, der „Advanced Placement AP Biology course“, der „introductory biology course for undergraduate majors“ von drei unterschiedlichen Universitäten, die „Biology Graduate Record Examination GRE” und mehrere „firstyear medical school courses from an institution with a traditional curriculum“). Es ist erfreulich, dass die schweizerischen Biologie-Maturaprüfungen mit einem entsprechenden Wert von 40.8% sich hier würdig in die grossen, aufwändig hergestellten Tests einreihen! Dieses Resultat ist aus drei Gründen nicht selbstverständlich:

a) Die US-amerikanische Pädagogik hat schon seit den 1960er Jahren damit begonnen, Fähigkeiten des „higher-order thinkings“ gezielt zu fördern, auch über die Art der Prüfungsfragen. Chipman, Segal & Glaser (1985) sowie Resnick (1987) geben einen (teilweise auch skeptischen) Rückblick auf die verschiedenen Trainingsprogramme wie beispielsweise das „productive thinking program“, „teaching thinking“ und „improving intelligence“. Die hohen Anteile von „higher-order questions“ in den aktuellen US-amerikanischen Assessments vierzig Jahre später erstaunen also nicht. Im deutschsprachigen Raum sind erst mit einer Verzögerung von zehn bis zwanzig Jahren einige Ansätze zur Förderung des produktiven Denkens (Kretschmer, 1981) oder sogenannte Denktrainings entwickelt worden (siehe dazu rückblickend Friedrich & Mandl, 1992, sowie Hager, 1995). Erfahrungsgemäss dauert es dann noch Jahre bis Jahrzehnte, bis sich selbst empirisch gut abgesicherte universitäre Überlegungen in der aktuellen Lehrerbildung niederschlagen. Es war also durchaus erwartbar, dass die konsequente Schulung neuer Lehrkräfte im Verfassen anspruchsvoller Prüfungsaufgaben im deutschsprachigen Raum erst vor einigen Jahren einsetzt hat und aktuell noch kaum geniessbare Früchte trägt. Die Ergebnisse von EVAMAR II weisen jedoch auf ein früheres Einsetzen dieses Wandlungsprozesses hin: die Resultate scheinen schon deutlich sichtbar zu sein.

b) Das Verfassen von guten Transferaufgaben benötigt sehr viel Zeit und auch ausserordentlich viel Fachwissen, das meist erst nach einer eigens für die Prüfungsaufgabe betriebenen Recherche vorliegt. Offenbar sind viele Lehrpersonen davon überzeugt, dass sich der Aufwand lohnt.

c) Die Korrektur von guten Transferaufgaben stellt hohe fachliche und zeitliche Anforderungen. Dies führt wahrscheinlich nicht selten zu einem Zielkonflikt der Lehrperson: Gemäss der Studie von Bol & Strage (1996) neigen Biologie-Lehrpersonen dazu, in Prüfungen eher wenig Transferaufgaben zu stellen, obwohl sie die Förderung von Transferleistungen als eines ihrer Unterrichtsziele ansehen. Auch scheint sich die Überzeugung durchzusetzen, dass sich eine aufwändige Korrektur lohnt.

Der hohe Prozentwert widerlegt auch das altbekannte Klischee, wonach in der Biologie durch blindes Auswendiglernen leicht eine gute Note zu erreichen und „Denken“ nicht nötig sei. Die Resultate lenken das Augenmerk aber auch auf die möglicherweise ungenügende Berücksichtigung der Themenbereiche Ökologie und Systematik/Evolution. Die Hinweise der Dozierenden zusammen mit dem relativ geringen notenrelevanten Gewicht dieser Themen dürften als Hinweis zu lesen sein, dass diese Gebiete im Biologieunterricht im Hinblick auf das Ziel der Studierfähigkeit zu oberflächlich behandelt werden und in den Prüfungen mehr Gewicht erhalten sollten.

Maturandinnen und Maturanden begegnen nicht einem ganz neuen Schwierigkeitsgrad von Aufgaben, wenn sie ein Studium ergreifen. Das zeigt der Vergleich zwischen den kognitiven Aktivierungspotenzialen, wie sie Maturandinnen und Maturanden von Aufgaben aus Maturitätsprüfungen kennen, und den kognitiven Aktivierungspotenzialen, denen sie in den Prüfungen des ersten Jahres an der Hochschule begegnen. Ein gutes Gedächtnis für das reiche Faktenwissen der Biologie (vgl. die beiden Anteile der Aufgabenkategorie 1.2 von fast 39% an der Hochschule und 47% am Gymnasium), wie es auch durch die „Auswendiglern-Aufgaben“ im Biologieunterricht

am Gymnasium trainiert wird, bleibt für ein gutes Abschneiden auch an Hochschulprüfungen immer noch eine solide Grundlage. Gross ist die Diskrepanz bei den Aufgaben der Kategorie 5.1, bei denen das präzise Erkennen von Fehlern in Fachaussagen verlangt ist. Multiple-Choice-Aufgaben mit als falsch oder richtig zu taxierenden Optionen sind für viele Studierende überraschend schwierig, was sich dadurch erklärt, dass das Erkennen von teilweise subtilen Fehlern in manchmal sehr dicht formulierten Fachaussagen an und für sich schwierig ist. Möglicherweise würde aber ein vermehrtes gezieltes Einüben dieses Aufgabentyps und dessen Einsatz in Maturitätsprüfungen die Studierfähigkeit unserer Gymnasiastinnen und Gymnasiasten noch verbessern.

VI Teilprojekt D2: Die Qualität von Maturaarbeiten in der Schweiz

Dieses Teilprojekt wurde vollständig durch die folgenden Autorinnen und Autoren der Pädagogischen Hochschule der Fachhochschule Nordwestschweiz konzipiert und durchgeführt: Christina Huber, Vera Husfeldt, Lukas Lehmann, Carsten Quesel (Leitung, E-mail:carsten.quesel@fhnw.ch). Unterstützt wurde dieses Team von Felix Boller (PH FHNW) und Eva Roos (IRDP).

Einleitung

Die in diesem Bericht im Zentrum stehende Beurteilung der Qualität der Maturaarbeiten soll hinsichtlich der im MAR 95 genannten allgemeinen Bildungsziele, insbesondere jedoch in Bezug auf überfachliche Qualifikationen, interdisziplinäre Aspekte und die wissenschaftspropädeutische Bildungswirkung erfolgen (vgl. Eberle, 2005). Ebenso sollen die Maturaarbeiten in Fortsetzung der Ergebnisse aus der ersten Evaluationsphase (EVAMAR I) vertieft evaluiert werden. Besonderes Augenmerk liegt dabei auf den Faktoren, welche die Qualität der Maturaarbeiten beeinflussen. Im Rahmen des Unterprojekts D2 wird zuerst zu klären sein, welche Merkmale die Qualität einer Maturaarbeit kennzeichnen. Erst wenn ein solcher Kriterienkatalog formuliert ist, kann die tatsächliche Auswertung der Arbeiten beginnen und schliesslich in eine Bewertung der Qualität der Maturaarbeiten in der Schweiz münden.

Nebst der eigentlichen Analyse der Maturaarbeiten sollen auch die institutionellen Rahmenbedingungen der Kantone und Gymnasien genauer betrachtet werden, die in Bezug auf Zielsetzung und Vorgehensweise beim Verfassen der Maturaarbeit massgebend sind. Eine dritte Analyseebene bilden die Maturandinnen und Maturanden. Sie werden sowohl zum Prozess des Verfassens der Maturaarbeit als auch zu Auswirkungen dieses Prozesses direkt befragt. Dabei stehen vor allem die Aspekte Themenwahl, Betreuung und Unterstützung, Ressourcen sowie Einschätzungen zu Aufwand und Nutzen der Maturaarbeit im Vordergrund. Abschliessend soll untersucht werden, welche Faktoren auf der Ebene der Einzelschulen respektive auf der Ebene der Maturandinnen und Maturanden einen Einfluss auf die Qualität der Maturaarbeiten haben.

Entsprechend diesen Überlegungen beinhaltet das vorliegende Projekt drei Komponenten. Zum einen wurden 920 Maturandinnen und Maturanden mittels eines standardisierten Fragebogens schriftlich befragt und insgesamt 437 Maturaarbeiten wurden anhand eines zuvor entwickelten, standardisierten Ratingbogens durch jeweils zwei unabhängige Personen beurteilt. Bei diesen beiden Teiluntersuchungen handelt es sich um quantitative Verfahren. Für die dritte Teiluntersuchung – die Analyse der kantonalen und schulischen Rahmenbedingungen – wurde ein qualitatives Vorgehen gewählt. Daran beteiligten sich insgesamt 30 Gymnasien aus der ganzen Schweiz.

Aus unterschiedlichen Gründen waren dieser Untersuchung Grenzen gesetzt. So war es im Rahmen dieses Projekts erstens nicht möglich, die Sichtweise der als Betreuerinnen und Betreuer fungierenden Lehrpersonen mit einem eigenen Instrumentarium zu untersuchen. Da es sich nicht um eine prozessbegleitende Untersuchung handelt, war es zweitens weder möglich, die pädagogische Interaktion im Kontext der Maturaarbeit zu analysieren, noch den durch die Projektarbeit erfolgten Kompetenzzuwachs auf Seiten der Schülerinnen und Schüler zu messen. Zudem konnte drittens die Beurteilungspraxis in den verschiedenen Sprachregionen nicht im Hinblick auf die Einheitlichkeit der Standards validiert werden, da für die ursprüngliche Absicht, im grösseren Umfang bilinguale Rater[36] einzusetzen, die entsprechenden sprachlichen und fachlichen Kompetenzen nicht in hinreichendem Masse zur Verfügung standen.

[36] Die englische Bezeichnung *Rater* wird hier geschlechtsneutral verwendet und bezieht sich mithin durchgängig sowohl auf männliche und weibliche Personen.

Aufbau des Berichtsteils

Im ersten Teil des Berichts werden einerseits die theoretischen Grundlagen sowie die für die Untersuchung zentralen Begriffe dargestellt und geklärt und der aktuelle empirische Forschungsstand zur Thematik der Maturaarbeiten dokumentiert (Abschnitt 1). Andererseits werden das Untersuchungsdesign sowie das konkrete methodische Vorgehen beschrieben (Abschnitt 2).

Im zweiten Teil werden die empirischen Forschungsergebnisse präsentiert: Abschnitt 3 widmet sich den kantonalen und schulischen Rahmenbedingungen. Der vierte Abschnitt hat die Resultate der Schülerinnen- und Schülerbefragung zum Gegenstand; im fünften Abschnitt werden schliesslich die Befunde der Ratings dargestellt und analysiert.

Danksagungen

Unser grösster Dank gebührt den Schülerinnen und Schülern, die sich die Zeit genommen haben, unseren Fragebogen auszufüllen und die uns ihre Maturaarbeiten zur Beurteilung zur Verfügung gestellt haben. Danken möchten wir aber auch allen beteiligten Schulen, und dort vor allem den Kontaktpersonen, die uns Auskunft gegeben und mit Material versorgt haben.

Herzlich gedankt sei des Weiteren allen Expertinnen und Experten, die den Aufwand auf sich genommen haben, sich im Umgang mit dem Kriterienraster schulen zu lassen und die Maturaarbeiten zu beurteilen.

Für die Unterstützung im Kanton Tessin bedanken wir uns bei Emanuele Berger (Direktor des Ufficio studi e ricerche [USR]) sowie bei Stefano Chiantese für die geleisteten Übersetzungsarbeiten. Für die Unterstützung in der französischsprachigen Schweiz danken wir den Kooperationspartnern vom Institut de recherche et de documentation pédagogique (IRDP) in Neuchâtel, namentlich Matthis Behrens als Direktor und Eva Roos als wissenschaftliche Mitarbeiterin, die uns bei der Entwicklung der Instrumente unterstützten sowie die Durchführung der Datenerhebung in der Romandie sicherstellten.

Eine bedeutende Rolle bei der Ausarbeitung des Instrumentariums spielte Felix Boller (PH FHNW), der aufgrund seiner langjährigen Erfahrung als Gymnasiallehrer und Fachdidaktiker ganz entscheidend zur Praxisnähe der Untersuchung beigetragen hat. Ihm sei hierfür ebenfalls ausdrücklich gedankt.

1 Ausgangslage und theoretischer Hintergrund

1.1 Die Maturaarbeit als Teil der Maturitätsreform 1995

1.1.1 Einführung der Maturaarbeit

Die Vorgeschichte der heutigen Maturaarbeit reicht mehrere Jahrzehnte zurück. So schlug die 1969 durch die EDK mandatierte Studienkommission in ihrem 1972 publizierten Bericht „Mittelschule von Morgen“ (MIMO) die Einführung schriftlicher Arbeiten vor (Meylan, 1996, S. 19f.). Mit der Einführung von „grössere[n] selbstständige[n] Arbeiten“, die integraler Bestandteil der Maturitätsprüfung sein sollten, strebte man eine individuelle Vertiefung von gymnasialen Lernprozessen und eine verbesserte Vorbereitung auf akademische Bildungsgänge an (EDK, 1972, S. 35). Die Arbeiten sollten „Selbstständigkeit im Denken, Erkennen des Wesentlichen, Logik der Gedankengänge, Sicherheit im Urteil, schöpferische Phantasie und Beherrschung der Fachmethodik“ (EDK, 1972, S. 37) fördern. Zudem sollte die Maturitätsprüfung durch die Einführung selbstständiger Arbeiten individueller gestaltet werden (EDK, 1972, S. 25). Der Entwurf sah vor, dass die Gymnasiastinnen und Gymnasiasten in zwei (von drei) Prüfungsfächern je eine selbstständige Arbeit ausführen und während des Maturitätsexamens je ein Examen in Form eines Prüfungsgesprächs ablegen.

Zu einer rechtsverbindlichen Aufwertung gymnasialer Projektarbeit kam es aber erst im Zuge der Revision der Maturitätsanerkennungsverordnung (MAV) in den 1990er Jahren. Im Maturitätsreglement (MAR) von 1995 wurde in Art. 10 festgelegt, dass Schülerinnen und Schüler „allein oder in einer Gruppe eine eigenständige schriftliche oder schriftlich kommentierte Arbeit erstellen und mündlich präsentieren“ müssen.

1.1.2 Idee der Maturaarbeit als interdisziplinäre Projektarbeit

Die wesentliche Idee hinter der Einführung der Maturaarbeit bestand darin, dass man den Aspekt der Interdisziplinarität sowie die Entwicklung von kognitiven, persönlichen und sozialen Schlüsselqualifikationen stärken wollte (EDK, 2000, S. 115f.; Meylan, 1996, S. 31f.). Die Maturaarbeit und ihre Doppelfunktion als Element des Kompetenzaufbaus und des Kompetenznachweises in der Schlussphase gymnasialer Bildungsgänge sind heute unumstritten. Bisweilen wird die Maturaarbeit gar als „das einzige echt innovative Element des MAR“ (Dreyer, 2005, S. 23) bezeichnet und zu den „unbestrittenen Highlights der Reform“ (Hirschi, 2002, S. 29) gezählt. Immer wieder betonen gymnasiale Lehrpersonen, dass sich die Maturaarbeit als Teil des gymnasialen Lernprozesses bewährt habe (vgl. Bonati, 2006, S. 6; Jacob, 2004, S. 5). Gleichwohl sprechen viele Indizien dafür, dass das Potential der Maturaarbeit als interdisziplinär gedachte Projektarbeit bislang nur unzureichend ausgeschöpft wird. Kritische Stimmen monieren, dass die Maturaarbeit in den meisten Fällen zu einer reinen Facharbeit verkommt (vgl. Schoch, 2003; Kunz, 2003). Dementsprechend wählten von den im Rahmen von EVAMAR I befragten Schülerinnen und Schüler lediglich 24% ein fächerübergreifendes Thema (Ramseier et al., 2005, S. 11).

Aktuelle politisch-administrative Entscheide der EDK und des Bundes bezeugen die Absicht, den Stellenwert der Interdisziplinarität in Zukunft zu erhöhen. Im Rahmen der letzten Teilrevision des MAR, die auf das Schuljahr 2008/09 hin in Kraft tritt, wurde in Artikel 11 explizit die Bestimmung getroffen, dass jede Schule sicher stellen soll, dass die Schülerinnen und Schüler „mit fächerübergreifenden Arbeitsweisen vertraut sind“.

1.1.3 Die Maturaarbeit im Prüfungskontext

Die Maturaarbeit zählt gemäss Art. 16 des MAR – mit Ausnahme des Kantons Luzern[37] – bisher nicht unmittelbar zu den Bestehensnormen der Maturitätsprüfungen. Allerdings wird die Bewertung der im Rahmen der Maturaarbeit erbrachten schriftlichen und mündlichen Leistungen unter Nennung des Themas im Maturitätsausweis aufgeführt (Art. 20 MAR) und in den meisten Kantonen ist die Zulassung zur Maturitätsprüfung davon abhängig, dass die Maturaarbeit als genügend bewertet wird. Im Fall des erfolgreichen Abschlusses erhalten die Schülerinnen und Schüler durch die Angabe von Titel und Benotung der Maturaarbeit im Zeugnis gleichsam eine "carte de visite", die sie mit mehr oder minder grossem Stolz vorweisen können (Fontolliet, 1996, S. 55).

Dessen ungeachtet fehlt es auf Seiten der Schülerinnen und Schüler nicht an kritischen Hinweisen, dass die Maturaarbeit für das Erreichen der Maturität zu wenig stark gewichtet werde (vgl. Binder & Feller-Länzlinger, 2004, S. 52; 2003, S. 22; Davaud & Hexel, 2003, S. 67; USO, 2006). So signalisierten 62% der im Rahmen von EVAMAR I befragten Schülerinnen und Schüler, dass die Bedeutung der Maturaarbeit für den Erwerb des Maturitätszeugnisses zu gering sei. Bemerkenswerterweise zeigten sich hier aber deutliche sprachregionale Unterschiede. So waren in der französischsprachigen Schweiz nur rund die Hälfte der Befragten (48%) dieser Ansicht, in der deutsch- und italienischsprachigen Schweiz dagegen waren es etwa zwei Drittel (D-CH: 63,4%, I-CH: 66,1%; vgl. Pagnossin, Alliata & Dozio, 2005, S. 190). Erklärungen für diese sprachregionalen Unterschiede liefert der Bericht zum ersten EVAMAR-Projekt jedoch nicht.

Demgegenüber zeigt sich in einer Befragung von Aargauer Gymnasiallehrpersonen, dass diese den Verzicht auf die prüfungsrelevante Anrechung der Maturaarbeit tendenziell eher begrüssen, auch wenn sie dem Anliegen der Schülerinnen und Schüler grundsätzlich folgen können. Die Lehrpersonen begründen ihre Zurückhaltung damit,

[37] Gemäss Reglement für die Maturitätsprüfungen im Kanton Luzern vom 27. Mai 1999 zählt die Maturaarbeit zu den Maturitätsfächern, die für das Bestehen der Maturitätsprüfung massgebend sind.

dass die Höhergewichtung eine Bewertung der Maturaarbeit erfordern würde, die verbindlichen und vergleichbaren Standards zu entsprechen vermag. Die uneingeschränkte Vielfalt an Themen und die Formen der Bearbeitung erschweren aber eine solcherart standardisierte Bewertung. Weiter merkten einige Lehrpersonen auch an, dass der fehlende Notendruck die intrinsische Motivation der Maturandinnen und Maturanden fördere und damit einen Gegenakzent zur Dominanz von primär auf die Noten fokussierten und insofern extrinsisch motivierten Leistungen darstelle (Binder & Feller-Länzlinger, 2004, S. 51f.).

Obwohl die Frage der Standardisierung der Bewertungsverfahren und Rahmenbedingungen der Maturaarbeit noch unbeantwortet ist, legt die im Jahr 2007 eingeleitete Teilrevision des MAR fest, dass die Maturaarbeit „aufgrund des Arbeitsprozesses der schriftlichen Arbeit und ihrer Präsentation" künftig zum Bestehen der Maturität zählt (vgl. EDK, 2007; Oelkers, 2008, S. 203ff.).

1.2 Die Maturaarbeit im Kontext gymnasialer Bildung

Traditionell war die gymnasiale Bildung durch die curriculare Fokussierung auf die Alten Sprachen und durch den funktionalen Bezug auf eine Universität geprägt, in der vor allem Theologie, Medizin und Jurisprudenz als akademische Kerndisziplinen gepflegt wurden. Im Laufe des 19. Jahrhunderts sah sich das altsprachliche Gymnasium einem ökonomisch-technisch induzierten Reformdruck ausgesetzt, der insbesondere auf die Aufwertung der Naturwissenschaften zielte. Das führte in vielen Ländern zur Ausdifferenzierung von Gymnasien mit geisteswissenschaftlichen und naturwissenschaftlich-technischen sowie später dann auch mit wirtschafts- und sozialwissenschaftlichen Profilen. Diese Differenzierung warf unweigerlich die Frage nach den Kriterien der Hochschulreife als Zugangsbedingung für das universitäre Studium auf. Trotz der lebhaften Diskussion um die Etablierung einer fachgebundenen Hochschulreife wurde letztlich an der allgemeinen Hochschulreife als höchstem Sekundarabschluss festgehalten.[38]

Die Binnendifferenzierung des Wissenschaftssystems im Laufe des 20. Jahrhunderts brachte nicht zuletzt auch eine gesteigerte Mehrdeutigkeit hinsichtlich der wissenschaftspropädeutischen Aufgaben der Gymnasien mit sich. Bestand früher eine starke Erwartung, dass sich die wissenschaftspropädeutische Qualität gymnasialer Bildungsprozesse relativ klar über den Bezug auf konzeptuelle Schnittmengen des Lehrbuchwissens einer übersichtlichen Zahl von Fachdisziplinen definieren lasse, musste die Frage der Wissenschaftspropädeutik infolge der Spezialisierung und Diversifikation akademischer Domänen neu gestellt werden. Diese Entwicklung führte in den letzten Jahrzehnten dazu, dass das Konstrukt der Allgemeinbildung in den gymnasialen Curricula einen neuen Stellenwert erhielt. Dies zeigt sich deutlich auch im Rahmenlehrplan für die Maturitätsschulen (EDK, 1994), in welchem festgehalten wird, dass das Gymnasium nicht Propädeutik für einzelne Wissenschaftszweige betreiben könne und dürfe, sondern dass den Schülerinnen und Schülern eine breite Grundlage an Kenntnissen, Fertigkeiten und Haltungen zu vermitteln sei, die es ihnen möglich mache, ihr Wissen auf jedem Gebiet und jederzeit zu erweitern (EDK, 1994, S. 6; vgl. Oelkers 2008, S. 74ff.).

Diese Idee des selbstgesteuerten Lernens wurde schon im frühen 20. Jahrhundert im Kontext der durch die „progressive education" entwickelten Projektmethode postuliert. Die Grundidee der Projektmethode lautet dahin, dass der Unterricht vom Kind her konzipiert werden muss, dass die Neugier und die Selbsttätigkeit der Kinder im Zentrum zu stehen habe und dass es beim Lernen vor allem darum gehe, sich aktiv mit Problemen auseinanderzusetzen und sich individuell oder kooperativ an deren Lösung zu versuchen (Frey, 1993; Oelkers, 1997). Innerhalb der reformpädagogischen Tradition fanden Recherchen und Experimente dabei bereits in der Primarschule einen festen Platz. Das eigenständige Sammeln und Auswerten von Informationen galt als ein Grundzug des schulischen Lernens, das wissenschaftliche Studium insofern als eine Fortsetzung des Einsatzes von kognitiven Strategien, die bereits in der Kindheit entwickelt werden.

[38] Vgl. in international vergleichender Perspektive Oelkers (2008, S. 185ff.), sowie speziell im Hinblick auf Deutschland Schäfer-Koch, (1997, S. 197f.).

Zunächst verstanden als ein Modell der Schulorganisation, hat sich die Projektmethode im Laufe des 20. Jahrhunderts zusehends zu einem flankierenden Moment innerhalb eines weitgehend unter fachdidaktischen Kriterien differenzierten Lehrplans entwickelt. Diese Beschränkung von Projekten verdankt sich vor allem der Einsicht, dass eine radikale Kindzentrierung letztlich einen vollständigen pädagogischen Steuerungsverzicht im Hinblick auf die Organisation und Strukturierung von Lernprozessen beinhalten würde (Oelkers, 1997, S. 27f.). In diesem Sinne stellt auch die Maturaarbeit ein ergänzendes Moment zum gymnasialen Curriculum dar. Aufgrund ihrer breiten Anforderungen ist sie – neben dem Unterricht in festgelegten Fächern – ein wesentliches Element in der Erfüllung des gymnasialen Bildungsauftrages (vgl. Binder, 2003; Eberle, 2007; Hadorn, 2004).

Der gymnasiale Bildungsauftrag kann heute als dreifacher Auftrag verstanden werden. So nennt die deutsche Kultusministerkonferenz (KMK) in der Vereinbarung zur Gestaltung der gymnasialen Oberstufe der Sekundarstufe II drei Bildungsziele des Gymnasiums: „Der Unterricht in der gymnasialen Oberstufe vermittelt eine vertiefte Allgemeinbildung, allgemeine Studierfähigkeit sowie wissenschaftspropädeutische Bildung“ (KMK, 2006, S. 5, eigene Hervorhebungen). Mit dem Aspekt der Vertiefung bei der Allgemeinbildung wird einerseits eine Kontinuität zu den Bildungszielen der Grundschule und der Sekundarstufe I hergestellt. Andererseits ist gerade die Wissenschaftspropädeutik auf den fortlaufenden Ausbau der Allgemeinbildung angewiesen, weil diese die „strenge Methode“ und die „formalisierte Arbeitsweise“ der wissenschaftlichen Disziplinen überschreitet und mit der Lebenswelt verbindet und damit auch ein transdisziplinäres Medium der Verständigung darstellt (von Hentig, 1996, S. 177). Bemerkenswert bei der Spezifikation der gymnasialen Bildungsziele durch die KMK ist, dass hier zwischen Studierfähigkeit und Wissenschaftspropädeutik differenziert wird. Diese Differenzierung wird im Hinblick auf den Unterricht näher bestimmt: Dieser führe die Schülerinnen und Schüler nicht nur „exemplarisch in wissenschaftliche Fragestellungen, Kategorien und Methoden ein“, sondern vermittle darüber hinaus „die Beherrschung eines fachlichen Grundlagenwissens“ (KMK, 2006, S. 5). Diese Beherrschung fachlichen Grundlagenwissens gilt als Schlüsselfaktor für die Verknüpfung von verschiedenen Wissensgebieten und für die Kultivierung einer selbstständigen und systematischen Arbeitsweise. Dieses Grundlagenwissen gilt zudem als förderliche Voraussetzung für die Entwicklung von Selbstverantwortung, Team- und Kommunikationsfähigkeit.

Der dreifache Bildungsauftrag, den die KMK für die deutschen Gymnasien formuliert, zeigt sich auch in Art. 5 des MAR zu den Bildungszielen: Absolventinnen und Absolventen der gymnasialen Ausbildung sollen „zu jener persönlichen Reife [gelangen], die Voraussetzung für ein Hochschulstudium ist und die sie auf anspruchsvolle Aufgaben in der Gesellschaft vorbereitet“ (MAR 95, Art. 5, Abs. 1). Im Folgenden gilt es, die in diesem relativ allgemein formulierten Bildungsauftrag enthaltenen Konstrukte der Studierfähigkeit oder Hochschulreife, der Wissenschaftspropädeutik sowie der überfachlichen Kompetenzen genauer zu definieren.

1.2.1 Studierfähigkeit und Hochschulreife

Andrä Wolter stellt die These auf, dass sich hinter dem Konstrukt der Studierfähigkeit ein unauflösliches und mehrdimensionales Heterogenitätsproblem (Wolter, 1997, S. 49) verbirgt, das seine Ursache einerseits in den bereits erwähnten Ausdifferenzierungsprozessen des Wissenschaftssystems, andererseits aber auch in gesamtgesellschaftlichen Entwicklungen (Pluralisierung der Informations- und Kommunikationswege, Pluralisierung der alltäglichen Lebenswelten und Sozialisationsfelder) hat. Diese Veränderungen führten sowohl in Bezug auf die Bildungsvoraussetzungen und die Erwartungen der Studierenden als auch bei den Studienanforderungen und den Erwartungen der Hochschulen zu einer Zunahme an Heterogenität (Wolter, 1997, S. 49). In empirischen Untersuchungen zeigt sich denn auch deutlich, dass die einzelnen Studienfächer vor allem in fachlicher und methodischer Hinsicht ganz unterschiedliche Erwartungen an ihre Studienanfängerinnen und -anfänger haben (vgl. Konegen-Grenier, 2002, S. 166; Heldmann, 1998). Einigkeit herrscht letztlich nur in Bezug auf die Tatsache, dass die Fähigkeiten unterschiedlichen Bereichen entspringen, nämlich personalen, formalen und materiellen Dimensionen (Konegen-Grenier, 2002). In diesem Sinne versteht Konegen-Grenier (2002, S. 79) Studierfähigkeit „als eine komplexe Fähigkeit [...], die sowohl eine kognitive, als auch eine persönliche, soziale und fachliche

Dimension umfasst.“ Aufgrund ihrer Befragung von Hochschuldozierenden kommt sie zu folgender inhaltlicher Bestimmung von Studierfähigkeit:

> „Analytische Fähigkeiten, Abstraktionsfähigkeit und Differenzierungsvermögen sind unabdingbare Voraussetzungen für ein Hochschulstudium. Sie müssen kombiniert sein mit guten Kenntnissen in den Fächern Englisch, Mathematik und Deutsch. Das Fachwissen ist durch Arbeitstechniken wie Präsentationsfähigkeit, Kenntnisse in der Textverarbeitung und Recherchetechniken zu ergänzen. Um seine kognitiven Fähigkeiten und sein fachliches Wissen anwenden zu können, muss der Studienanfänger eine persönliche Arbeitshaltung mitbringen, die inhaltliches Interesse mit Leistungsbereitschaft und Genauigkeit vereint“ (Konegen-Grenier, 2002, S. 168).

Diese inhaltliche Bestimmung des Konstrukts der Studierfähigkeit deutet eine enge Verknüpfung, aber auch eine fehlende Trennschärfe in Bezug auf die Konstrukte der Wissenschaftspropädeutik und der überfachlichen Kompetenzen an.

Im Rahmen von EVAMAR I wurden Maturandinnen und Maturanden um eine Einschätzung bezüglich ihrer Vorbereitung auf ein Hochschulstudium befragt. Dabei zeigte sich, dass die generelle Einschätzung dieser Vorbereitung positiv ausfällt. 76% der Maturandinnen und Maturanden fühlen sich gut oder eher gut vorbereitet (Ramseier, Allraum & Stalder, 2005, S. 127). In Bezug auf einzelne Studienrichtungen zeigte sich jedoch ein differenzierteres Bild. Je nach Studienrichtung variierte der Anteil derjenigen Maturandinnen und Maturanden, die sich als gut oder eher gut vorbereitet fühlten, zwischen 80% (für das Studienfach Geschichte) und 14% (für Ingenieurwissenschaften). Neben Geschichte wurde vor allem die Vorbereitung auf Sprachwissenschaften, Naturwissenschaften und Mathematik oft als (eher) gut bezeichnet. Der Umstand, dass die Vorbereitung auf die Ingenieurwissenschaften am schwächsten abschneidet, ist in erster Linie darauf zurückzuführen, dass es sich hier um eine Domäne handelt, die in vielfacher Hinsicht quer zum Kanon der Schulfächer steht, so dass den Schülerinnen und Schülern hier von der Unterrichtspraxis her keine klaren Antizipationen entwickeln können (Ramseier, Allraum & Stalder, 2005, S. 128). EVAMAR I zeigte ausserdem, dass die Vorbereitung auf das tatsächlich gewählte Hochschulstudium positiv ausfällt: 54% der Maturandinnen und Maturanden reagierten mit „stimmt eher“ und 22% mit „stimmt genau“ auf die Aussage „Ich bin gut auf das Studium vorbereitet, das ich gewählt habe“ (Ramseier, Allraum & Stalder, 2005, S. 133).

Etwas weniger positive Resultate zeigt die TOSCA-Studie, die für Abiturientinnen und Abiturienten in Baden-Württemberg im Längsschnittvergleich ergeben hat, dass sie die Qualität der Studienvorbereitung im Gymnasium noch verhalten positiv bewerten, nach Aufnahme eines Studiums aber dann retrospektiv zu skeptischeren Bewertungen gelangen. Im Durchschnitt stellen sie ihren Schulen zwar kein schlechtes Zeugnis aus, bringen aber auch kein ausgeprägtes Vertrauen in die erworbene Hochschulreife zum Ausdruck (Trautwein et al., 2006, S. 56f).

1.2.2 Wissenschaftspropädeutik

Im MAR finden sich lediglich zwei knappe Hinweise auf die Funktion des Gymnasiums als wissenschaftspropädeutische Einrichtung: In Artikel 5 zu den Bildungszielen ist einerseits die Rede von „persönlicher Reife“ als „Voraussetzung für das Hochschulstudium“ und andererseits vom Ziel der „Einsicht in die Methodik wissenschaftlichen Arbeitens“. Darin zeigt sich auch die Nähe zum Begriff der Hochschulreife oder Studierfähigkeit. Andererseits macht die Formulierung auch deutlich, dass das Konstrukt der Wissenschaftspropädeutik kaum pädagogisch operationalisiert werden kann und dass sich die Wissenschaftspropädeutik heute stärker auf die Arbeitsmethoden und weniger auf die Inhalte bezieht (vgl. Criblez, 2005, S. 5). Dementsprechend ist die Wissenschaftspropädeutik auch im Rahmen der TOSCA-Studie in Baden-Württemberg in erster Linie anhand von methodologischen Gesichtspunkten operationalisiert worden. Dazu gehört etwa die Literaturrecherche, das Schreiben von Protokollen und das Durchführen von Experimenten (Trautwein & Lüdtke, 2004, S. 362f.).

Der methodologische Fokus zeigt sich auch deutlich bei dem im angloamerikanischen Sprachraum verwendeten Begriff der „Scientific Literacy“. Die National Academy of Science hat eine Definition für diese Basiskompetenz

in Anwendung gebracht, die sich auf das Verständnis der Regeln wissenschaftlicher und technischer Rationalität konzentriert.

> "Scientific literacy means that a person can ask, find, or determine answers to questions derived from curiosity about everyday experiences. It means that a person has the ability to describe, explain, and predict natural phenomena. Scientific literacy entails being able to read with understanding articles about science in the popular press and to engage in social conversation about the validity of the conclusions. Scientific literacy implies that a person can identify scientific issues underlying national and local decisions and express positions that are scientifically and technologically informed. A literate citizen should be able to evaluate the quality of scientific information on the basis of its source and the methods used to generate it. Scientific literacy also implies the capacity to pose and evaluate arguments based on evidence and to apply conclusions from such arguments appropriately." (National Academy of Science, 1996)

Geht man von dieser Definition der Scientific Literacy aus und formuliert sie allgemeiner ohne die Berücksichtigung der Logik der naturwissenschaftlichen Disziplinen, so kann man einen Begriff der Wissenschaftsorientierung und Wissenschaftspropädeutik daraus ableiten, der auf die spezifische Kompetenz der Schülerinnen und Schüler hinweist, verschiedene fachliche Verfahren zu beherrschen und zu verstehen.[39] Ein solches, nicht an die Logik bestimmter Disziplinen gebundenes Konzept kommt der „Science Literacy", die von Bonati als typische gymnasiale Kompetenz verstanden wird, sehr nahe. Zentrale Elemente wissenschaftlicher Methoden können auf der Grundlage dieser Kompetenz in vereinfachter Form verarbeitet werden (Bonati, 2007, S. 51). So verstanden meinen Wissenschaftsorientierung und Wissenschaftspropädeutik nichts anderes, als dass die Schülerinnen und Schüler lernen, elementare Techniken des methodisch kontrollierten Erkenntnisgewinns einzusetzen und dadurch Evidenzen zu erlangen, die eine systematische Urteilsbildung ermöglichen.

Aufgrund der Ausdifferenzierung des Wissenschaftssystems und der damit verbundenen Spezialisierungsprozesse ist es schwierig, einen für alle Fachdisziplinen geltenden Kanon von Grundregeln aufzustellen. Trotzdem lassen sich allgemeine Anforderungen an die Wissenschaftspropädeutik stichwortartig umreissen: Nach Schmidt (1994, S. 226ff.) besteht die wichtigste Leistung der Wissenschaftspropädeutik darin, dass sie zu Methodenbewusstsein und Methodenbewusstheit führt, zumal wissenschaftliche Erkenntnisse stets durch methodisches Verhalten gewonnene Erkenntnisse sind. Für die gymnasiale Bildung bedeutet dies, dass Gymnasiastinnen und Gymnasiasten Kenntnisse über wesentliche Strukturen und Methoden der verschiedenen Wissenschaften gewinnen, also in wissenschaftliche Verfahrensweisen eingeführt werden und dabei lernen, mit diesen reflektierend umzugehen. Dabei sollen sie auch die Fähigkeit und Bereitschaft entwickeln, die eigene Vorgehensweise erkenntnis- und wissenschaftstheoretisch zu reflektieren, mit anderen Ansätzen zu konfrontieren und in den Kontext relevanter Forschungstraditionen einzuordnen. Gymnasiastinnen und Gymnasiasten sollen auch die Grenzen wissenschaftlicher Aussagen erkennen und verstehen können, dass der Erwerb von Erkenntnissen, der stets von Interessen abhängig ist, dem Wandel und der Fortentwicklung unterworfen ist. Schliesslich kann auch die Fähigkeit, theoretisches Wissen zu kommunizieren und anzuwenden, mit zur Wissenschaftspropädeutik gezählt werden (vgl. Huber, 1998, S. 161; KMK, 1978, S. 561; Köller et al., 2004, S. 117; Schmidt, 1994, S. 226ff.).

Mit Huber (1997, S. 348, Hervorhebungen i.O.) kann die Wissenschaftspropädeutik in drei Ebenen differenziert werden:

1. Lernen und Einüben in Wissenschaft, d.h. Grundbegriffe und -methoden kennen lernen;
2. Lernen an Wissenschaft, d.h. eine Haltung des Immer-weiter-fragens, Begründens und Abwägens einnehmen können;
3. Lernen über Wissenschaft, d.h. kritische Reflexion in grösseren Zusammenhängen.

[39] Im Unterschied zur Academic Literacy, die sich weitgehend auf die Entwicklung der Lese- und Schreibkompetenz beschränkt (vgl. Neeley, 2005).

Huber (1998, S. 150) macht auch darauf aufmerksam, dass Wissenschaftspropädeutik ohne Spezialisierung und vertiefte Arbeit an einer Frage- oder Problemstellung nicht denkbar ist. So gesehen kann die Wissenschaftsorientierung des Unterrichts als Basis für den Umgang mit Expertenwissen verstanden werden (Tenorth, 1994, S. 176). Die Wissenschaftspropädeutik beinhaltet darüber hinaus die ersten Schritte zum Aufbau von Expertenwissen.

Aus der Zielformulierung im MAR wird klar, dass Wissenschaftspropädeutik keineswegs ausschliesslich als Studienvorbereitung gesehen werden kann, sondern dass ihr Anspruch – gerade im gymnasialen Kontext – sich auch auf die Bildung der Persönlichkeit erstreckt (vgl. Schmidt, 1994, S. 223), womit die Schnittmenge zum Bereich der überfachlichen Kompetenzen ersichtlich wird.

1.2.3 Überfachliche Kompetenzen

Die Frage, was genau unter überfachlichen Kompetenzen verstanden werden soll, findet bis anhin keine abschliessende Antwort (Notter & Arnold, 2003, S. 12); in der Fachliteratur trifft man auf diverse Kataloge und Auflistungen zu diesem Topos. So wurden etwa im Rahmen der Studie „Objectifs pédagogiques transversaux“ (EVAMAR I, Teilprojekt 2) vier fächerübergreifende Kompetenzbereiche unterschieden (Grin et al., 2004, S. 40):

1. Kulturkritische Kompetenzen, d.h. die Fähigkeit, ideologische Standpunkte aufzudecken und die eigene Voreingenommenheit erkennen zu können, Offenheit gegenüber anderen Kulturen bezeugen und die eigene Kultur unter einem anderen Blickwinkel betrachten sowie unterschiedliche Formen künstlerischen Schaffens schätzen zu können.
2. Informationsverarbeitungskompetenzen, d.h. wissen, wo und mit welchen Hilfsmitteln Informationen gewonnen werden können, verschiedene Medien angemessen zur Informationsbeschaffung nutzen zu können und Kritikfähigkeit in Bezug auf die Anwendung neuer Technologien zu zeigen.
3. Wissenschaftliche Kompetenzen, d.h. die Fähigkeit, wissenschaftlich argumentieren, theoretische Aussagen formulieren, Hypothesen aufstellen und überprüfen zu können sowie in der Lage zu sein, die eigene Ausdrucksweise dem Kontext anzupassen und sich klar und präzise auszudrücken.
4. Kooperative Kompetenzen, d.h. sich selbst und anderen gegenüber Verantwortung übernehmen zu können sowie selbstständig und in Gruppen arbeiten zu können.

Bieri Buschor und Forrer (2005, S. 42) definieren überfachliche Kompetenzen ganz allgemein als „Potenzial einer Person, in unterschiedlichen Situationen und in unterschiedlichen sozialen Rollen angemessen und verantwortungsbewusst zu handeln“. Dabei erstreckt sich der Aspekt des überfachlichen Lernens darauf, dass die Kompetenzen nicht speziell in einzelnen Schulfächern erworben werden. Das bedeutet nicht, dass überfachliche Kompetenzen generell das fachbezogene Lernen ersetzen. Ganz im Gegenteil: „Die Forschung legt sogar nahe, dass die Entwicklung fächerübergreifender Kompetenzen das Vorhandensein gut ausgeprägter fachbezogener Kompetenzen voraussetzt“ (Klieme et al., 2003, S. 75). Im Rahmen von EVAMAR I zeigte sich, dass Lehrpersonen, die Förderung von überfachlichen Kompetenzen befürworten und davon ausgehen, dass die Förderung fächerübergreifender Kompetenzen die persönliche Entwicklung der Schülerinnen und Schüler anregt und zudem einen Beitrag im Hinblick auf ein späteres Studium oder die Berufsbildung leistet (Ramseier et al., 2005, S. 13).

Notter & Arnold (2003, S. 12) befragten Studierende zu verschiedenen überfachlichen Kompetenzen zum Zeitpunkt ihrer Maturität. Im Rahmen ihrer Untersuchung zeigte sich, dass die Studierenden die Kompetenzen Textverständnis (Verstehen und Interpretieren), selbstständiges Arbeiten, schriftliche Ausdrucksfähigkeit und selbstständiges Lernen zum Zeitpunkt ihrer Matur als am besten ausgebildet einschätzen. Am schlechtesten wurden die Kompetenzen Auseinandersetzen mit ethischen Fragen, Auftreten vor Publikum und Zeiteinteilung eingeschätzt (Notter & Arnold, 2003, S. 23). Es zeigte sich ausserdem, dass Studierende aller Fachrichtungen alle überfachlichen Kompetenzen als relativ wichtig einstuften. Als besonders wichtig wurden die folgenden Kompetenzen eingestuft:

- selbstständiges Lernen;
- selbstständiges Arbeiten;

- Verantwortung für das eigene Lernen übernehmen;
- Zeiteinteilung;
- Umgehen mit Belastungen;
- Problemlösefähigkeit: Probleme erkennen, Lösungen suchen und erarbeiten (Notter & Arnold, 2003, S. 39).

Bieri Buschor und Forrer (2005) haben überfachliche Kompetenzen bei jungen Erwachsenen in der Schweiz aufgrund von Befragungsdaten analysiert. Der Anspruch ihrer Studie geht dahin, Aussagen darüber machen zu können, wie gut die jungen Erwachsenen sich gerüstet fühlen, allgemeinen Anforderungen postadoleszenter Lebensgestaltung im öffentlichen Raum und in der Privatsphäre gerecht werden zu können. Die Systematik der Kompetenzbereiche erstreckt sich von den Selbstkompetenzen über das Kooperations- und Lernvermögen bis hin zur sozialen und ökologischen Verantwortung sowie der politischen Kompetenz. Dabei ist recht schnell erkennbar, dass die Klassifikation der Kompetenzbereiche grosse Schnittmengen aufweist: So erweist es sich etwa analytisch als schwierig, einen Trennstrich zwischen der Selbstreflexion und den kognitiven Lernstrategien zu ziehen.

1.2.4 Zwischenbilanz

Mit der Einführung der Maturaarbeit verbindet sich die Erwartung, dass Schülerinnen und Schüler des Gymnasiums heute veränderten Anforderungen bei der Flexibilität und der Selbstorganisation gerecht werden müssen. Bei der Maturaarbeit handelt es sich dem Anspruch nach zum einen um ein Medium der Aneignung solcher Kompetenzen, zum anderen aber auch um einen Leistungsnachweis, der den erfolgreichen Erwerb dieser Kompetenzen dokumentieren soll. Hinsichtlich der Eingrenzung der pädagogischen Funktion der Maturaarbeit zeigt sich, dass bis anhin die verschiedenen Konzepte der Hochschulreife und der Wissenschaftspropädeutik ebenso wie die der überfachlichen Kompetenzen uneinheitlich spezifiziert werden. Als schwacher Konsens zeichnet sich ab, dass die Schülerinnen und Schüler im Prozess des Erstellens der Maturaarbeit individuell oder im Team über einen längeren Zeitraum projektbezogen arbeiten, dass sie sich vertieft mit ausgewählten Methoden des Recherchierens, Analysierens und Argumentierens vertraut machen und sich in der Lage zeigen, elementare wissenschaftliche Informationen sachgerecht zu verarbeiten. Die Erwartungen an die Maturaarbeiten erstrecken sich nicht nur auf rezeptive und wissensgenerierende Leistungen, sondern auch auf die Performanz: Das im Prozess erworbene Wissen muss – in Form eines schriftlichen Berichts und einer mündlichen Präsentation – angemessen kommuniziert werden. Mithin soll die Maturaarbeit als wesentliches Element bei der Genese der Hochschulreife und Studierfähigkeit nicht nur eine wissenschaftspropädeutische Funktion im engeren Sinne erfüllen, sondern auch zum Ausbau von überfachlichen Kompetenzen beitragen, in deren Zentrum die eigenverantwortliche und souveräne Bewältigung von Aufgaben steht.

1.3 Forschungsstand zur Thematik der Maturaarbeiten

Einige wenige Studien haben sich bereits mit Aspekten der Thematik um die Maturaarbeit befasst: In der ersten Phase von EVAMAR wurden im Rahmen des Teilprojekts 2 zu den fächerübergreifenden Zielsetzungen ca. 3'500 Lehrpersonen und rund 13'000 Schülerinnen und Schüler, die sich im letzten Jahr ihrer Maturitätsausbildung befanden, mittels einer standardisierten Erhebung befragt (Pagnossin et al., 2005). Weiterhin liegt eine vergleichende Darstellung der gesetzlichen und reglementarischen Rahmenbedingungen der gymnasialen Maturaarbeit und der interdisziplinären Projektarbeit der Berufsmaturität vor (Zillig, 2004). Ausserdem gibt es für die Kantone Aargau (Binder & Feller-Länzlinger, 2004), Genf (Davaud & Hexel, 2003), Luzern (Binder & Feller-Länzlinger, 2003) und Waadt (Stocker, 2006) Evaluationsberichte zur Umsetzung des MAR 95, in denen Aussagen zur Maturaarbeit gemacht werden. Und für ein Freiburger Gymnasium liegen zwei Artikel über eine schulinterne Untersuchung zur Maturaarbeit vor (Darbellay, 2002; Schmid, 2002). Im Folgenden werden die wichtigsten Erkenntnisse aus diesen Studien zusammengefasst.

1.3.1 Typen von Maturaarbeiten

Im MAR 95 werden keine Aussagen zur inhaltlichen Ausrichtung oder zur Form der Maturaarbeit gemacht. Die Studie von Pagnossin et al. (2005) gibt jedoch erste Hinweise aus der Praxis. Es lassen sich demnach fünf grundsätzliche Typen von Maturaarbeiten unterscheiden:

- theoretische Arbeiten (41%);
- empirische Arbeiten (33%);
- künstlerische, audiovisuelle und materielle Produktionen (12%);
- ausserschulische Aktivitäten (12%) und
- soziokulturelle Projekte (2%).

Zu betonen ist, dass diese Klassifikation nicht trennscharf ist: So lässt sich etwa geltend machen, dass künstlerische Produktionen durchaus eine empirische Komponente aufweisen.

In der Verteilung der Arbeitstypen zeigen sich deutliche Unterschiede zwischen den Sprachregionen. Empirische Arbeiten sind in der Deutschschweiz am beliebtesten, in der italienisch- und französischsprachigen Schweiz dagegen werden theoretische Arbeiten vorgezogen. Künstlerische und audiovisuelle Maturaarbeiten werden in der Deutschschweiz signifikant häufiger eingereicht als in der lateinischen Schweiz. In Bezug auf das Geschlecht der Maturandinnen und Maturanden zeigen sich nur zwei signifikante Unterschiede: Männer neigen eher zu theoretischen Arbeiten als Frauen (m: 42,5%, f: 39%) und Frauen verfolgen häufiger soziokulturelle oder ausserschulische Projekte als Männer (m: 11,5%, f: 16%). Die Autoren dieses Teilprojekts von EVAMAR führen die sprachregionalen Unterschiede auf die unterschiedlichen kantonalen Bestimmungen zurück. Die geschlechtsspezifischen Unterschiede werden aufgrund ihrer geringen Ausprägung nicht weiter erklärt (Pagnossin et al., 2005, S. 185f.).

1.3.2 Themenwahl

Zillig (2004, S. 15f.) stellt in ihrer Untersuchung fest, dass die Praxis der Themenwahl an den einzelnen Schulen verschieden ausgestaltet wird. Grundsätzlich ist die Wahlfreiheit in der deutschsprachigen Schweiz grösser als in der französisch- oder italienischsprachigen Schweiz. Dieses Bild bestätigt sich auch in der Untersuchung von Pagnossin et al. (2005): 97% der befragten Schülerinnen und Schüler der deutschsprachigen Schweiz wählten ihr Thema frei. Von den Befragten der französischsprachigen Schweiz waren 65% bei der Themenwahl insofern eingeschränkt, als sie aus einem vorgegebenen Themenbereich oder einer Themenliste wählen mussten. Bei den italienischsprachigen Schülerinnen und Schülern konnten nur gerade 8% ihr Thema frei wählen. Dies ist darauf zurückzuführen, dass die Betreuung der Schülerinnen und Schüler im Kanton Tessin in der Form von Seminaren gewährleistet wird, die themenspezifisch angeboten werden (Pagnossin et al., 2005, S. 184f.).

Die Themenwahl erfolgte bei einer grossen Mehrheit der Befragten interessengeleitet. Es wurden jedoch auch andere Faktoren wie Vorkenntnisse, Drittpersonen oder der Faktor „Herausforderung“ genannt (Binder & Feller-Länzlinger, 2003, S. 9; 2004, S. 41; Pagnossin et al., 2005, S. 184f.). Die Untersuchung an einem deutschsprachigen Gymnasium in Fribourg zeigt, dass 61% der befragten Schülerinnen und Schüler mit der Themenfestlegung einverstanden waren (Schmid, 2002, S. 11).

Von den im Rahmen von EVAMAR I befragten Schülerinnen und Schüler wählten nur gerade 24% ein Thema, das mehr als ein Schulfach betraf, 39% verankerten ihre Maturaarbeit in nur einem Schulfach und 38% verfassten die Arbeit in einer nicht-schulischen Disziplin (Pagnossin et al., 2005, S. 186). Zu diesem Resultat muss angemerkt werden, dass die Rahmenbedingungen an den einzelnen Schulen einen grossen Einfluss haben können, zumal die Wahl von Themen aus dem ausserunterrichtlichen Bereich nicht an allen Schulen erlaubt wird (Zillig, 2004, S. 15f.).

1.3.3 Durchführung

Eine grosse Mehrheit der Schülerinnen und Schüler verfassen ihre Maturaarbeit alleine (Pagnossin et al., 2005, S. 191; Binder & Feller-Länzlinger, 2003). Eine Ausnahme bilden die befragten Maturandinnen und Maturanden des Kantons Aargau: 76% von ihnen erstellten ihre Maturaarbeit im Team. Binder und Feller-Länzlinger erklären diesen Unterschied einerseits mit den Vorgaben des Kantons Aargau, in denen betont wird, dass Maturaarbeiten als Gruppenarbeit durchgeführt werden sollen, und andererseits damit, dass auch die vorgängig durchgeführte Projektarbeit in der Regel im Team erfolgt (Binder & Feller-Länzlinger, 2004, S. 42f.). Die Autoren erfragten in ihren Untersuchungen in den Kantone Luzern und Aargau auch den Zeitaufwand, den die Schülerinnen und Schüler in die Erarbeitung der Maturaarbeit investierten. Dabei zeigte sich, dass das Ausmass der investierten Zeit sehr stark nach Gymnasium variiert und dass im Kantonalvergleich die Luzerner Maturandinnen und Maturanden tendenziell mehr Zeit in die Maturaarbeit investierten als ihre Aargauer Kolleginnen und Kollegen. Die Ursache für diese Differenz vermuten Binder und Feller-Länzlinger in der Tatsache, dass die Aargauer Schülerinnen und Schüler im Team arbeiten und folglich individuell je weniger Zeit investieren müssen.

Die Maturaarbeiten werden in den meisten Schulen in einem Zeitraum von zehn bis 18 Monaten verfasst (Zillig, 2004, S. 17). Auf die Frage, wie viel Zeit zwischen dem Start- (Projektvertrag, Arbeitsvereinbarung) und Abgabetermin für die Erarbeitung der Maturaarbeit zur Verfügung stehen soll, nannten die meisten der befragten Luzerner und Aargauer Schülerinnen und Schüler eine Zeitspanne von sechs Monaten (Binder & Feller-Länzlinger, 2003, S. 12; 2004, S. 44). In der schulinternen Umfrage eines Freiburger Gymnasiums wurde eine Zeitspanne von rund einem Jahr als zu lang beurteilt. Die grosse Mehrheit (63%) der befragten Lehrpersonen, Schülerinnen und Schüler erachteten jedoch auch an dieser Schule weniger als 8 Monate als ausreichend (Schmid, 2002, S. 10).

Diverse Schulen stellen für die Maturaarbeit Unterrichtszeit zur Verfügung, z.B. im Stundenplan eingebaute Lektionen, Blockwochen etc. (Zillig, 2004, S. 18). Nach solchen Zeitgefässen befragt, wünschten sich 28% der Luzerner Maturandinnen und Maturanden Unterrichtszeit im Rahmen von einem bis fünf Tagen, 32% von sechs bis zehn Tagen und 19% von 11 bis 15 Tagen zur Bearbeitung der Maturaarbeit (Binder & Feller-Länzlinger, 2003, S. 14).

1.3.4 Wissen, Fähigkeiten und Fertigkeiten der Schülerinnen und Schüler

Zwei Drittel der im Rahmen von EVAMAR I befragten Lehrpersonen waren der Ansicht, dass die Fähigkeit der Schülerinnen und Schüler, selbstständig eine Arbeit zu verfassen, ausreichend sei (Pagnossin et al., 2005, S. 177). Bei den Schülerinnen und Schülern waren es 91%, die ihre eigene Fähigkeit zum selbstständigen Arbeiten als ausreichend einschätzten (Pagnossin et al., 2005, S. 188). In Bezug auf das methodische Vorwissen der Schülerinnen und Schüler war nur rund die Hälfte der befragten Lehrpersonen der Überzeugung, dass diesbezüglich hinlängliche Kompetenzen vorhanden waren. Bei den Schülerinnen und Schülern waren dagegen sieben von zehn der Meinung, dass sie vorgängig genügend methodologische Kenntnisse erworben hätten, um selbstständig eine Arbeit verfassen zu können (Pagnossin et al., 2005, S. 177; S. 188).

Unzufriedener äusserten sich Gymnasiastinnen und Gymnasiasten in der Freiburger Studie. Von ihnen empfanden rund die Hälfte im französischsprachigen Teil der Schule und 58% im deutschsprachigen Teil die Einführung ins wissenschaftliche Arbeiten als unzureichend (Darbellay, 2002, S. 14; Schmid, 2002, S. 11).

Bei der Evaluation der MAR 95-Umsetzung im Kanton Aargau zeigte sich, dass der zur Vorbereitung auf die Maturaarbeit durchgeführte Projektunterricht von den Schülerinnen und Schülern eher kritisch beurteilt wird. Die im Rahmen dieses Projektunterrichts vermittelten Fertigkeiten seien im Hinblick auf das Verfassen der Maturaarbeit als nur beschränkt nützlich einzustufen (Binder & Feller-Länzlinger, 2004, S. 45).

1.3.5 Betreuung

Der Entstehungsprozess der Maturaarbeit wird an allen Schulen begleitet, die Betreuungspraxis ist aber nicht einheitlich (Zillig, 2004, S. 17). Die Mehrheit der Luzerner Maturandinnen und Maturanden (80%) traf sich zwischen ein bis acht Mal mit der betreuenden Lehrperson. Bei diesen Treffen ging es vorrangig um inhaltliche Tipps, Methodenfragen des wissenschaftliches Arbeiten, die Themenwahl, Präsentation der Arbeit und Verfassen des Berichts (Binder & Feller-Länzlinger, 2003, S. 17f.). An der schulinternen Umfrage am deutschsprachigen Gymnasium in Fribourg äusserten 47% der Befragten, dass die Begleitung der Maturaarbeit vermehrt auf Gesprächen aufbauen sollte. Zudem zeigte sich, dass sich etwa ein Drittel der Schülerinnen und Schüler durch die Begleitung eingeengt fühlten, 13% jedoch hätten mehr Anweisungen erwartet (Schmid, 2002, S. 10).

In der gesamtschweizerischen Untersuchung von Pagnossin et al. (2005, S. 177ff.) äusserten nur 54% der befragten Lehrpersonen, dass sie mit der Zeit, die ihnen für die Betreuung zur Verfügung gestellt wird, zufrieden sind. 43% waren der Ansicht, dass die ihnen zur Verfügung gestellten Zeitressourcen für die Betreuung ausreichen. Die Erfahrung am deutschsprachigen Gymnasium in Fribourg zeigte, dass eine Entlastung der Lehrperson von einer Stunde auf acht Seminarteilnehmerinnen und -teilnehmer nicht ausreicht: „Damit Aufwand und Ertrag etwa übereinstimmen, wäre eine Entlastungsstunde pro höchstens 5 bis 6 Schülerinnen und Schüler notwendig“ (Schmid, 2002, S. 11).

Interessanterweise zeigt sich auf Schülerinnen- und Schülerseite – zumindest im Kanton Waadt – ein anderes Bild. Dort sind nämlich drei Viertel der befragten Schülerinnen und Schüler zufrieden oder sehr zufrieden mit der für die Betreuung zur Verfügung stehenden Zeit (Stocker, 2006, S. 35). Ebenso zeigte sich auch in EVAMAR I, dass die Mehrheit der Schülerinnen und Schüler die Betreuung – trotz der aus pädagogischer Sicht knappen Zeitressourcen – als ausreichend beurteilten (Pagnossin et al., 2005, S. 177ff.). Diese Zufriedenheit bestätigte sich ebenso in der Luzerner Studie (Binder & Feller-Länzlinger, 2003, S. 16).

Grundsätzlich erachtet die Mehrheit der Schülerinnen und Schüler wie auch der Lehrpersonen die zur Verfügung gestellten materiellen Ressourcen als ausreichend. Es fällt in der Studie von Pagnossin et al. jedoch auf, dass die Meinungen zwischen Lehrpersonen sowie Schülerinnen und Schülern beim Thema der Rahmenbedingungen und Ziele divergieren: Während die meisten Lehrpersonen meinen, diese seien hinreichend klar festgelegt, beurteilten nur 55,6% der befragten Schülerinnen und Schüler die ihnen zur Verfügung gestellten Wegleitungen als ausreichend (Pagnossin et al., 2005, S. 177ff.). Ein ähnliches Resultat zeigte sich auch bei den Gymnasiastinnen und Gymnasiasten in den Kantonen Waadt und Genf. In der Waadtländer Untersuchung gaben 45% der Befragten an, dass die geltenden Richtlinien und Weisungen nicht klar genug seien (vgl. Davaud & Hexel, 2003, S. 63; Stocker, 2006, S. 35). Für die Genfer Gymnasien stellten Davaud und Hexel auch fest, dass in Gymnasien, in denen die Schülerinnen und Schüler mit den Realisierungsbedingungen eher unzufrieden sind, auch der Nutzen der Maturaarbeit eher in Frage gestellt wird (Davaud & Hexel, 2003, S. 62ff.).

1.3.6 Zwischenbilanz

Die hier dargestellten Studien liefern erste wichtige Befunde zu den Maturaarbeiten. Gleichwohl bleiben viele Fragen offen: Die Resultate der Studie von Pagnossin et al. (2005) beruhen nur auf Einschätzungen der befragten Lehrpersonen und Schülerinnen und Schülern. Da die Arbeiten selbst nicht zum Gegenstand der Untersuchung gemacht wurden, lassen sich hier zu Qualitätsmerkmalen allenfalls indirekte Aussagen treffen. Die Studien von Binder & Feller-Länzlinger (2003, 2004), Davaud & Hexel (2003) und Stocker (2006) erheben zwar auch Daten zu den Maturaarbeiten, doch handelt es sich bei ihren Evaluationsberichten um kantonale Untersuchungen, die keine verallgemeinerten Aussagen auf nationaler Ebene erlauben; noch enger ist der Blickwinkel der schulinternen Untersuchung aus Fribourg (Darbellay, 2002; Schmid, 2002). Die Studie von Zillig (2004) zeigt auf, dass die Rahmenbedingungen kantonal und im Schulvergleich variieren, doch wird nicht untersucht, ob und inwiefern diese unterschiedlichen Rahmenbedingungen Auswirkungen auf die Qualität der Maturaarbeiten haben.

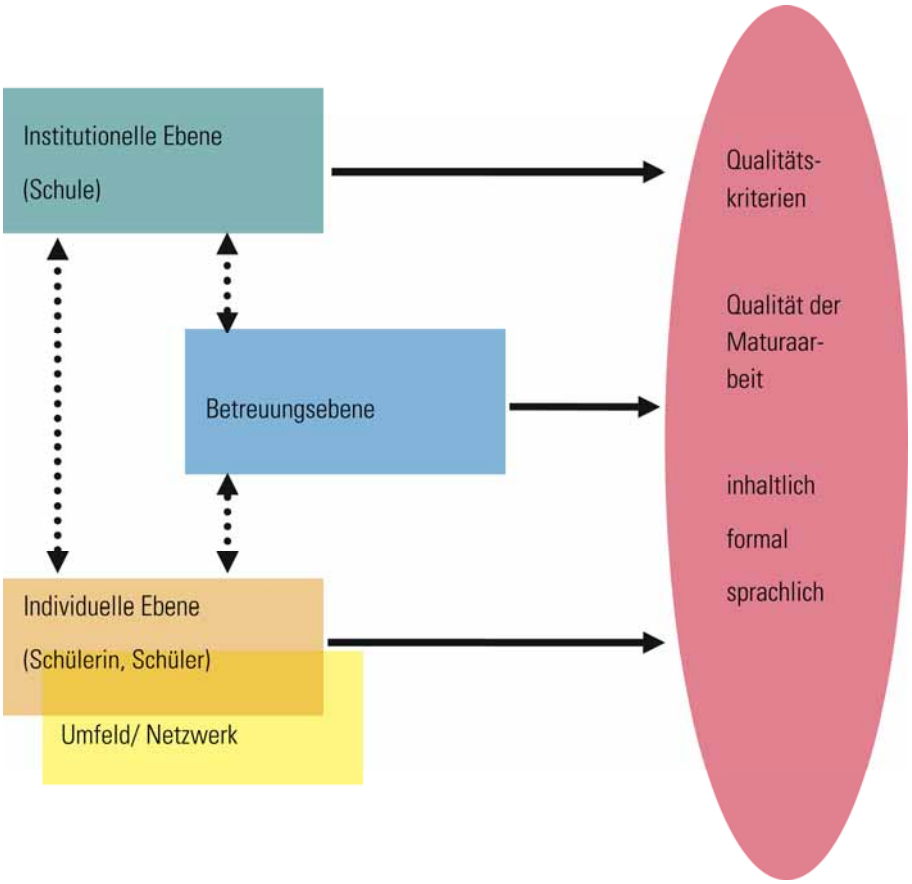
2 Methodische Grundlagen und Durchführung der Untersuchung

Im Folgenden wird zunächst das Untersuchungsdesign für die Analyse der Qualität der Maturaarbeitens dargestellt. Den Ausgangspunkt bildet ein Wirkungsmodell, das den "Produktionsprozess" der Maturaarbeiten erfasst. Des Weiteren geht es um die Entwicklung der Erhebungsinstrumente sowie das konkrete Vorgehen bei der Datenauswertung. Ebenso werden die Stichprobenziehung und die untersuchte Grundgesamtheit dargestellt.

2.1 Wirkungsmodell

Zur Planung der Untersuchung und zur Gewinnung von Hypothesen wurde ein Wirkungsmodell entwickelt (vgl. Abbildung VI.1), das die für die Analyse der Maturaarbeiten relevanten Zusammenhänge abbildet.

Abb. VI.1: Wirkungsmodell zur Genese der Maturaarbeiten



Auf der Grundlage dieses Wirkungsmodells ist die Untersuchung in drei Teilkomponenten untergliedert worden. Im Zentrum steht dabei die Analyse eines nationalen Samples von Maturaarbeiten anhand eines Kriterienrasters. Flankierend sind zudem die institutionellen Rahmenbedingungen, in welche die Gymnasiastinnen und Gymnasiasten im Prozess des Erstellens einer Maturaarbeit eingebunden sind, in einer komparativen Perspektive untersucht worden. Des Weiteren wurden alle Schülerinnen und Schüler der ausgewählten Klassen mittels eines standardisierten Fragebogens schriftlich befragt. Dabei interessierten neben soziodemographischen Daten und allgemeinen Angaben zur Maturaarbeit (Thema und Themenwahl, Bewertung der Arbeit durch Lehrperson) vor allem die Betreuungssituation (Form und Häufigkeit der Betreuung, Zufriedenheit mit der Betreuung) und Aspekte der Unterstützung und Ressourcen (Zufriedenheit mit den schulischen Rahmenbedingungen, Unterstützungsangebote, soziales Netzwerk). Ausserdem wurden die Schülerinnen und Schüler um eine Einschätzung ihres individuellen Aufwands wie auch zum Nutzen der Maturaarbeit gebeten.

2.2 Beurteilung der Maturaarbeiten

2.2.1 Zuschnitt der Untersuchung

Im Fokus der Untersuchung steht die Maturaarbeit als Produkt eines Arbeitsprozesses, der auf Grundlage der vorliegenden Daten nicht im Detail rekonstruiert werden kann. Dies hat zur Folge, dass psycho- und soziometrische Aspekte in der Beurteilung der Qualität der Maturaarbeiten weithin ausser Betracht bleiben müssen. Ebenso bleibt die tatsächliche Eigenleistung der Gymnasiastinnen und Gymnasiasten unbekannt, da sich nicht eruieren lässt, wie viel Unterstützung sie beispielsweise von familiärer Seite oder aus dem Bekanntenkreis erhalten haben. Auf diese Problematik macht auch Oswald Inglin, Konrektor des Gymnasiums Leonhard (Basel), aufmerksam, wenn er die Maturaarbeit als „familiäres Joint Venture" bezeichnet: „Ganze Heerscharen von Eltern, Geschwistern und Freunden verbringen eine Unmenge Zeit mit Durchlesen und Veredeln von Maturaarbeiten. Je nach familiärem Hintergrund sind gewisse Schülerinnen und Schüler in dieser Hinsicht bevorteilt, respektive Maturandinnen und Maturanden aus so genannt bildungsferneren Familien sind benachteiligt" (Inglin, 2007, S. 27).

Die Einschätzung der Qualität einer Maturaarbeit lässt zudem keinen gesicherten Schluss auf die Kompetenzen der Verfasserin oder des Verfassers der Arbeit zu. Beim Projekt Maturaarbeit handelt es sich um einen einmaligen Arbeitsprozess, dessen Nachhaltigkeit nicht unmittelbar am Produkt abgelesen werden kann: Es ist nicht gewährleistet, dass die zum Ausdruck gebrachten Kompetenzen bereits verfestigt sind und auch in einem anderen Kontext (z.B. Hochschulstudium) und unter leicht veränderten Bedingungen (z.B. weniger intensiver Betreuung) erfolgreich genutzt werden können (vgl. Bieri Buschor & Forrer, 2005, S. 44). Diese Fokussierung auf die Maturaarbeit als Produkt eines Arbeitsprozesses hat jedoch den entscheidenden Vorteil, dass ein Halo-Effekt vermieden werden kann, da der Schüler oder die Schülerin der beurteilenden Person unbekannt bleibt.

2.2.2 Bezugsnormen zur Beurteilung der Qualität von Maturaarbeiten

Die Qualität von Maturaarbeiten lässt sich nur anhand einer Bezugsnorm bestimmen, die als Bewertungsmassstab dient. Für das hier gewählte Untersuchungsdesign ist die Orientierung an inhaltlich beschriebenen Standards am plausibelsten. Eine individuelle Bezugsnorm kommt nicht in Frage, da das Produkt eines Prozesses beurteilt wird, ohne den Prozess selbst beobachten zu können. Soziale Bezugsnormen werden in der vorliegenden Untersuchung zwar eine Rolle spielen, doch weisen sie blinde Flecken auf, die nicht vernachlässigt werden dürfen: Erstens ist ein Vergleich aufgrund sozialer Merkmale wie etwa Sprachregion, Geschlecht usw. nur innerhalb unserer Stichprobe möglich. Da unsere Stichprobe aber – gerade für den französischsprachigen Teil der Schweiz – Verzerrungen aufweist (vgl. Kap. 2.6 unten), besteht die Gefahr, dass hier Fehleinschätzungen entstehen (vgl. Rheinberg, 2002, S. 64ff.). Zweitens setzt auch ein Vergleich innerhalb unserer Stichprobe voraus, dass Einschätzungen zur Qualität der Maturaarbeiten bestehen, d.h. dass vorgängig eine sachlich-inhaltliche Bezugsnorm gesetzt wurde.

Bisher fehlt es an Instrumenten zur Beurteilung von Maturaarbeiten anhand sachlicher Bezugsnormen. Zwar finden sich in der Fachliteratur Kompetenzmodelle und Kataloge von Kompetenzen, die im Rahmen des Verfassens schriftlicher Arbeiten erworben oder weiterentwickelt werden. Oftmals sind diese Kataloge aber wenig systematisch oder wirken willkürlich (vgl. etwa Bonati & Hadorn, 2007; Wilhelmer, 2005, S. 85; Winter, 2006) und sind deshalb für unser Forschungsvorhaben wenig geeignet.

Hinweise auf die sachlich-inhaltliche Bezugsnorm zur Beurteilung von Maturaarbeiten sind unter Umständen in den Anforderungen ersichtlich, die von verschiedenen Seiten an Maturaarbeiten und ihre Verfasserinnen und Verfasser gestellt werden. Hierzu gehören:

1. Anforderungen aus Sicht der gymnasialen Bildungsinstitutionen (gymnasiale Bildungsziele),
2. Anforderungen aus Sicht der Abnehmenden, d.h. der Hochschulen aber auch der Berufsbildung (Studierfähigkeit, Schlüsselqualifikationen),
3. Anforderungen aus Sicht der Disziplinen (Wissenschaftlichkeit),
4. Anforderungen aus Sicht der Linguistik.

ad 1) Anforderungen aus Sicht der gymnasialen Bildungsinstitutionen

Die Bildungsziele der gymnasialen Bildungsinstitutionen, werden im MAR 95 (EDK, 1995) und im Rahmenlehrplan der Maturitätsschulen (EDK, 1994) beschrieben. In diesen beiden Dokumenten finden sich keine konkreten Aussagen zur Maturaarbeit oder zu den Zielen, welche im Prozess des Verfassens einer Maturaarbeit erreicht werden sollen. Die beschriebenen Bildungsziele liefern aber durchaus Orientierungshilfen, welche als allgemeine Leitplanken zur Formulierung inhaltlicher Bezugsnormen dienen können (vgl. hierzu Kap. 1.2 oben und Kap. 3.1.1 unten).

Im Rahmen von EVAMAR I wurde untersucht, welchen Qualitätsaspekten bzw. Bewertungskriterien Gymnasiallehrpersonen sowie Gymnasiastinnen und Gymnasiasten besonders hohe Wichtigkeit zuweisen. Die Bewertungskriterien wurden in fünf Bereiche eingeteilt:

1. wissenschaftliche Vorgehensweise,
2. Eigenleistung der Studierenden,
3. Präsentation der Arbeit,
4. interdisziplinärer Charakter der Arbeit,
5. Neuartigkeit/Originalität der Arbeit.

Es zeigte sich, dass grundsätzlich fast alle Aspekte als wichtig erachtet wurden. So gaben bei den drei Bereichen Wissenschaftliche Vorgehensweise, Eigenleistung der Studierenden und Präsentation der Arbeit jeweils mehr als 80% der befragten Lehrpersonen an, dass sie diese als wichtig erachten. Der interdisziplinäre Charakter der Arbeit sowie die Originalität oder die Neuartigkeit einer Arbeit dagegen wurden weniger häufig als wichtig erachtet (Interdisziplinarität <50%, Originalität 53,5%). Auch auf Seiten der Schülerinnen und Schüler wurden die Eigenleistung sowie die Präsentation der Arbeit als wichtig erachtet. Weniger oft stuften sie aber die wissenschaftliche Vorgehensweise als wichtig ein (vgl. Pagnossin et al., 2005, S. 179f.).

Da im Rahmen der hier durchgeführten Untersuchung nur die Endprodukte, also die Maturaarbeiten als Textdokument vorliegen, können die als wichtig bewerteten Aspekte Eigenleistung der Studierenden und Präsentation der Arbeit nicht berücksichtigt werden. Aber der von den Lehrpersonen als wichtig erachtete Aspekt der wissenschaftlichen Vorgehensweise, der in der Formulierung der „Einsicht in die Methodik wissenschaftlichen Arbeitens“ auch in Artikel 5 des MAR 95 formuliert wird, wurde in der vorliegenden Untersuchung als Bezugsnorm aufgenommen (vgl. Kapitel 2.2.3).

Den Anforderungen gymnasialer Bildungsinstitutionen wurde zusätzlich Rechnung getragen, indem verschiedene, zufällig ausgewählte Bewertungsraster einzelner Schulen synoptisch dargestellt wurden. Diese Synopse diente letztlich als Grundlage für die Entwicklung des verwendeten Beurteilungsinstrumentes.

ad 2) Anforderungen aus Sicht der Abnehmenden

Aus Sicht der Abnehmenden scheinen vor allem die Aspekte der Schlüsselqualifikationen und der überfachlichen Kompetenzen zentral zu sein (vgl. auch Kap. 1.2). Da es – wie bereits weiter oben ausgeführt – nicht möglich ist, aufgrund des Produktes eines einmaligen Arbeitsprozesses auf tatsächlich vorhandene Kompetenzen zurückzuschliessen, kann dieser Aspekt innerhalb eines Querschnittdesigns nicht analysiert werden. Die erfolgreiche Erarbeitung und Präsentation der Maturaarbeit stellt im Hinblick auf die Studierfähigkeit ein notwendiges, aber kein hinreichendes Kriterium dar. Dieses Kriterium bleibt insofern unscharf, weil die Hochschulen nicht auf eine allgemein gültige inhaltliche Definition der Studierfähigkeit bauen können. Jedoch stellen akademische Ausbildungseinrichtungen ähnliche Anforderungen an die Schreibkompetenz, den sorgfältigen Umgang mit Daten und Fakten, die Einhaltung von methodologischen Regeln, das selbstständige analytische Denken und die Schlüssigkeit des Argumentierens, die in Richtlinien für die Gestaltung von schriftlichen Studienarbeiten geltend gemacht werden (vgl. exemplarisch Erdfelder, Musch & Cüpper, 2002; Carolus, 2006; Wallmeier, 2007; des Weiteren Konegen-Grier, 2002).

ad 3) Anforderungen aus Sicht der Disziplinen

Beim ersten Bezugspunkt (Gymnasien als abgebende Institutionen) hat sich gezeigt, dass Gymnasiallehrpersonen den Aspekt der wissenschaftlichen Vorgehensweise als besonders wichtig erachten. Beim zweiten Bezugspunkt (Hochschulen als aufnehmende Institutionen) fungiert die Wissenschaftlichkeit als institutionelles Grundmerkmal. Die einzelnen Disziplinen haben aber je eigene Ansprüche an die Theoriebildung und den Methodengebrauch und stützen sich auf unterschiedliche Traditionen der Professionalisierung. Für die Analyse der Maturaarbeiten ist es deshalb geboten, jede Arbeit durch mindestens eine Fachexpertin oder einen Fachexperten beurteilen zu lassen. Dieses Verfahren stellt sicher, dass der Aspekt der Wissenschaftlichkeit anhand der jeweiligen disziplinären Standards bewertet werden kann.

ad 4) Anforderungen aus Sicht der Linguistik

Im Bereich der Linguistik existieren verschiedene Instrumente zur Analyse von Texten, so etwa das „Zürcher Textanalyseraster“ (Nussbaumer & Sieber, 1994). Diese Instrumente fokussieren vor allem auf linguistische Aspekte eines Textes und eignen sich deshalb nur bedingt zur Beurteilung von Maturaarbeiten. Das „Zürcher Textanalyseraster“, das „die Feststellung von Qualitäten und Mängeln in Texten und damit ein (re)konstruierendes Verstehen der Kohärenz von Texten“ (Nussbaumer & Sieber, 1994, S. 185) ermöglicht, lieferte aber wertvolle Hinweise zur Entwicklung des Rasters zur Analyse von Maturaarbeiten.

Das Zürcher Textanalyseraster besteht aus drei Teilen:

- Null-Teil: Bezugsgrössen/Korrelate.

Dieser Teil misst vor allem Eigenschaften eines Textes, so beispielsweise die Textlänge, den Umfang und die Varianz des Vokabulars und der syntaktischen Gestaltung, den Aufwand an und die Art der expliziten Kohäsionsleistung sowie die grundsätzliche Schwierigkeit des im Text behandelten Themas (Nussbaumer & Sieber, 1994, S. 156f.).

- A-Teil: Sprachsystematische und orthographische Richtigkeit.

Hier werden die drei Aspekte Graphie (Orthographie und Interpunktion), Grammatik (Morphologie, Syntax und Textbau) sowie Semantik untersucht; mithin geht es um die Erfassung der sprachformalen Korrektheit eines Textes (Nussbaumer & Sieber, 1994, S. 160f.)

- B-Teil: Angemessenheit des Sprachmittelgebrauchs.

Im Unterschied zum A-Teil wird hier nicht dichotomisch in richtig oder falsch unterteilt, vielmehr erfolgt die Einschätzung auf einer Skala von besonders angemessen bis besonders unangemessen (Nussbaumer & Sieber, 1994, S. 161ff.).

Bei der Analyse der Maturaarbeiten wurde explizit auf die Untersuchung von Texteigenschaften im Sinne des Null-Teils des Zürcher Textanalyserasters verzichtet. Denn das Ziel der Untersuchung ist es, zu einer qualitativen Aussage über die untersuchten Maturaarbeiten zu kommen und nicht quantitative Aspekte der schriftsprachlichen Artikulation zu erfassen.

Die Aspekte der sprachsystematischen und orthographischen Richtigkeit sowie die Angemessenheit des Sprachmittelgebrauchs (vgl. A- und B-Teil des Zürcher Textanalyserasters) jedoch spielen eine Rolle in Bezug auf die Qualität von Maturaarbeiten. Da die Untersuchung aber v.a. sozialwissenschaftlicher und weniger linguistischer Natur ist, wurden sprachformale Aspekte nur anhand der übergeordneten Kriterien Text (Textfluss, Sprachliche Prägnanz, Wortwahl/Wortschatz, Ausdruck, Satzstrukturen) und Sprachliche Korrektheit untersucht.

Nussbaumer und Sieber (1994, S. 167) machen darauf aufmerksam, dass gewisse Kriterien besondere Schwierigkeiten in Bezug auf die Objektivierbarkeit hervorrufen und dass die Beurteilungen von Texten in Bezug auf diese Kriterien (sie nennen etwa die Qualität der verwendeten sprachlichen Mittel) zwischen den beurteilenden Personen massiv divergieren können. Diese Schwierigkeit zeigt sich auch in der Untersuchung der Qualität der Maturaarbeiten. Dass die Expertinnen und Experten im Rahmen dieser Evaluation in wechselnden Dyaden und in unterschiedlichen Anteilen beteiligt waren, erschwert die Feststellung der Interrater-Reliabilität ausserordentlich: Als Prüfkriterium für die Konsistenz des Urteilsverhaltens greifen wir deshalb auf die Intraklassenkorrelation beim Globalurteil zurück.

Weitere linguistische Anhaltspunkte wurden anhand der sprachpragmatischen Überlegung gewonnen, dass sich an verschiedene Textarten unterschiedliche funktionale Anforderungen stellen. Kruse (1999) unterscheidet vier Anforderungen an Texte:

1. Genauigkeit: Unter Genauigkeit wird die Präzision verstanden, mit der ein Text die Intention des Verfassers oder der Verfasserin zum Ausdruck bringt.
2. Kommunikative Funktion: Die kommunikative Funktion eines Textes ergibt sich aus der sozialen Wirkung, welche ein Text erfüllen soll.
3. Strukturiertheit: Sie bezieht sich auf den inneren Aufbau des Textes und um die Verknüpfung seiner Elemente.
4. Emotionale Authentizität: Hierunter wird das Ausmass gefasst, in welchem der expressive Aspekt eines Textes das emotionale Erleben der Textproduzentin oder des Textproduzenten widerspiegelt.

Der vierte Aspekt der emotionalen Authentizität spielt bei Maturaarbeiten eine eher untergeordnete Rolle. Die anderen drei Aspekte aber sind in Bezug auf Maturaarbeiten durchaus von Bedeutung. Ausgehend von Kruses Anforderungen an Texte, wurde das Kriterienraster in die drei Anforderungsbereiche Kommunikative Funktion, Strukturiertheit und Genauigkeit gegliedert.

2.2.3 Entwicklung des Analyseinstrumentes

Auf der Grundlage der Überlegungen zu Bezugspunkten der Beurteilung von Maturaarbeiten wurde das Kriterienraster in einem zirkulären, sowohl induktiven als auch deduktiven Prozess entwickelt, mehrfach getestet und überarbeitet, bis es angemessen und praktikabel erschien. Die mehrfache Prüfung des Rasters auf seine Anwendbarkeit und Tauglichkeit hin führte nicht nur zu einer Optimierung des Instruments, sondern war auch mit dem Nebeneffekt verbunden, dass die an der Entwicklung beteiligten Personen in der Anwendung des Rasters „geeicht“ wurden.

Die Operationalisierung des Rasters erfolgte letztlich auf einer sechsstufigen Skala. Die ursprüngliche Absicht, eine vierstufige Skala zu verwenden und jeden einzelnen Skalenpunkt zu operationalisieren, erwies sich als problematisch, zumal es nicht möglich war, die Beschreibungen der einzelnen Niveaus ausreichend trennscharf und präzise vorzunehmen. Abgesehen von der Schwierigkeit, alle Bewertungsstufen zu versprachlichen, erwies sich die vierfache Unterteilung in der Erprobung des Instruments als zu grob, um die Qualitätsunterschiede der Arbeiten angemessen abzubilden. Aufgrund der Vorüberlegung, dass keine neutrale mittlere Antwortkategorie angeboten werden sollte, erfolgte ein Wechsel zu einer sechsstufigen Skala, bei der jeweils die beiden Extrempole für jedes Item verbalisiert worden sind.

2.2.4 Ratingverfahren

Jede Maturaarbeit wurde durch zwei Rater begutachtet. Hierbei wurde darauf geachtet, dass immer mindestens eine der beiden Personen über Fachkompetenzen in derjenigen Disziplin verfügen muss, in welcher die Maturaarbeit verfasst wurde. Bei interdisziplinären Arbeiten wurde darauf geachtet, dass möglichst alle tangierten Fachbereiche abgedeckt werden konnten. In der Regel handelte es sich bei den Fachexpertinnen und -experten um Personen mit einem abgeschlossenen Hochschulstudium in einem oder mehreren der in der Arbeit tangier-

ten Disziplinen, in den meisten Fällen verfügten diese Fachleute zudem über ausgewiesene Erfahrung im Bereich der Fachdidaktik auf der Sekundarstufe II. Für die Zweitratings wurden zudem Personen eingesetzt, die über erziehungs- und sozialwissenschaftliche Qualifikationen, über mehrjährige Forschungserfahrung und in der Regel auch über Kompetenzen in der akademischen Lehre verfügen.

Die Erfahrungen im Rahmen des Entwicklungsprozesses des Analyserasters zeigten, dass eine Schulung der Rater unabdingbar ist, um eine möglichst hohe Interrater-Reliabilität und damit Objektivität der Aussagen zu gewinnen. Im Zuge der Schulung wurde den Ratern jeweils ein Handbuch zur Verwendung des Kriterienrasters übergeben.

Die Beurteilung der Maturaarbeiten erfolgte anonymisiert, damit ausgeschlossen werden konnte, dass sich die Rater von anderen Faktoren wie etwa dem Geschlecht der Schülerinnen und Schüler beeinflussen lassen.

2.2.5 Überprüfung von Plagiaten

Zur Beurteilung, ob es sich bei den Arbeiten oder bei Teilen um Plagiate handelt, wurden die Expertinnen und Experten aufgefordert eine Einschätzung dazu abzugeben. Der von einem Rater ausgesprochene Verdacht beweist freilich noch nicht, dass tatsächlich ein Plagiat vorliegt, andersherum ist das Gegenteil nicht bewiesen, wenn kein Verdacht ausgesprochen wird. Aus diesem Grund wurden aus der Gruppe der verdächtigen sowie der unverdächtigen Fälle zufällig je zehn Arbeiten gezogen, die einer vertieften Überprüfung unterzogen wurden.

Von den 20 Arbeiten, die vertieft danach analysiert wurden, ob plagiiert wurde, lagen 13 in elektronischer Form vor, so dass sie einem Test mittels des Computerprogramms Plagiarism Finder unterzogen werden konnten.[40] Am Ende der Auswertung zeigt Plagiarism Finder jeweils eine Statistik an, in der unter anderem aufgeführt wird, wie gross der Prozentsatz kongruenter Wörter ist, welcher im Internet gefunden wurde. Ausserdem werden die verdächtigen Stellen direkt im Dokument angezeigt und die entsprechenden Links angegeben, so dass dann an Referenztexten überprüft werden kann, ob es sich tatsächlich um ein Plagiat handelt oder nicht. In einem zweiten Schritt wurden alle 20 zufällig ausgewählten Arbeiten analog des von Weber-Wulff (o.J.) vorgeschlagenen Verfahrens mittels "Google-Technik" überprüft, d.h. auffällige Textstellen wurden in Google eingetippt und so überprüft, ob allenfalls von Internetseiten abgeschrieben wurde.

[40] Bei Plagiarism Finder handelt es sich um eine Software, die den Text analysiert und einzelne Wörter separiert. In einem zweiten Schritt werden die Wörter jeweils im Zusammenhang (also beispielsweise sieben zusammenhängende Wörter) in einer Internetsuchmaschine abgefragt.

2.3 Vergleichende Darstellung der Rahmenbedingungen

Abb. VI.2: Rahmenbedingungen im Wirkungsmodell

☞ Wirkungsmodell

	[→	Betreuung	→]	
Institutionelle Ebene	→	↕	→	Qualität der Maturaarbeit
	[→	Schüler, Schülerin (Individuelle Ebene)	→]	

Ausgehend von dem in Abbildung 1 dargestellten Modell kann die institutionelle Ebene sowohl einen direkten als auch einen indirekten Einfluss auf die Qualität der Maturaarbeiten haben. Grundsätzlich werden nationale, kantonale sowie einzelschulische Vorgaben unterschieden.

Auf nationaler Ebene wurden einerseits das MAR 95 und andererseits der Rahmenlehrplan analysiert. Im Gegensatz zum MAR 95 sind die Gymnasien nicht verpflichtet, den Rahmenlehrplan zu berücksichtigen, zumal es sich bei diesem um eine rein formelle Empfehlung zuhanden aller Kantone handelt, ihre gymnasialen Lehrpläne aufgrund dieser Vorlage zu gestalten.

Kantonale Bestimmungen zur Funktion der Maturaarbeiten sind auf unterschiedlichen gesetzlichen Ebenen festgehalten. Aus diesem Grunde werden für die Analyse der Rahmenbedingungen systematisch alle kantonalen Gesetze und Verordnungen danach untersucht, ob sie Hinweise auf die Maturaarbeit enthalten. Zudem wurden auch die kantonalen Lehrpläne im Hinblick auf die Thematik untersucht.

Die Darstellung der Rahmenbedingungen auf der Ebene der Einzelschulen basiert auf der Inhaltsanalyse der durch die Schulen schriftlich dokumentierten Informationen (Weisungen, Reglemente, Leitfäden usw.), welche im Schuljahr 2006/07 Gültigkeit besassen, da sich das Sample auf die Vorgaben dieses Zeitraums bezieht. Die von den Schulen zur Verfügung gestellten Informationen wurden anonymisiert.

2.4 Befragung der Schülerinnen und Schüler

Abb. VI.3: Schülerinnen und Schüler im Wirkungsmodell

☞ Wirkungsmodell

Schüler, Schülerin (Individuelle Ebene)		
↕	→	Qualität der Maturaarbeit
Betreuungsebene		

Im Rahmen der dritten Teilerhebung wurde ein standardisierter Fragebogen für die Schülerinnen und Schüler entwickelt, der folgende Aspekte umfasst:

- soziodemographische Variablen,
- Angaben zur Maturaarbeit (Fachliche Einordnung der Arbeit, Themenwahlprozess, Bewertung der Arbeit),
- Betreuung (Form und Häufigkeit der Betreuung, Zufriedenheit mit der Betreuung),
- Unterstützung / Ressourcen (Unterstützungs-/Vorbereitungsangebote, Zufriedenheit mit den schulischen Rahmenbedingungen),
- Einschätzung von Aufwand und Nutzen der Arbeit.

Bei der Entwicklung des Fragebogens wurde auf die im Rahmen von EVAMAR I durchgeführte Befragung der Abschlussklassen zurückgegriffen (vgl. Ramseier et al., o.J.). Die Mehrheit der Fragen, die damals speziell zur Maturaarbeit gestellt wurden, sind – teilweise in leicht modifizierter Form – übernommen worden, um eine längsschnittliche Perspektive andeuten zu können. Da es nicht möglich ist, anhand der Maturaarbeiten Aussagen darüber zu gewinnen, ob diese zu Fortschritten in Bezug auf die im MAR genannten Bildungsziele geführt hat, wurden die Schülerinnen und Schüler ausserdem um eine Einschätzung darüber gebeten, inwiefern die Maturaarbeit zur Erreichung dieser Ziele beigetragen hat.

2.5 Überprüfung des postulierten Wirkungsgefüges

Nach Abschluss der drei Teilerhebungen werden die Daten miteinander kombiniert, damit untersucht werden kann, inwiefern die im Wirkungsmodell angenommenen Zusammenhänge tatsächlich nachweisbar sind. Die Hypothesenbildung erstreckt sich dabei auf die vier im Wirkungsmodell genannten Bereiche:

1. Qualität der Maturaarbeit
2. Institutionelle Ebene:
   a) Themenwahl
   b) Vorbereitung
   c) Realisierungsbedingungen
3. Betreuungsebene
4. Individuelle Ebene:
   a) Motivation und Selbstwahrnehmung
   b) Intrinsischer und extrinsischer Wert der Arbeit.

Die Hypothesenbildung stützt sich hinsichtlich der Qualität der Maturaarbeiten auf den oben dargestelltten Diskurs zur Entwicklung überfachlicher Kompetenzen; dasselbe gilt für die Aspekte auf der individuellen Ebene. Hinsichtlich der institutionellen Ebene und der Betreuung stützt sich die Hypothesenbildung auf den Diskurs zu den Bildungszielen des Gymnasiums.

2.5.1 Qualität der Maturaarbeit

In der vorliegenden Untersuchung gehen wir davon aus, dass sich die Qualität der Maturaarbeiten sinnvoll in die drei Bereiche der inhaltlichen, der formalen und der sprachlichen Qualität unterteilen lässt und dass zwischen diesen Qualitätsbereichen bedeutsame Zusammenhänge bestehen, die sich in der Beurteilung widerspiegeln (→ Hypothese 1a). Des Weiteren gehen wir davon aus, dass es bei den Maturaarbeiten hinsichtlich der drei genannten Qualitätsaspekte geschlechtsspezifische, sprachregionale und arbeitsformspezifische (Einzelarbeit vs. Gruppenarbeit) Unterschiede gibt (→ Hypothese 1b), wobei die jeweilige disziplinäre oder interdisziplinäre Zuordnung keine Rolle spielt (→ Hypothesen 1c, 1d).

Tab. VI.1: Hypothesen zu übergreifenden Qualitätsaspekten

1	Übergreifende Qualitätsaspekte
1a	Es gibt eine positive Korrelation zwischen den inhaltlichen, formalen und sprachlichen Qualitätsaspekten der Maturaarbeiten.
1b	Es gibt signifikante geschlechtsspezifische, sprachregionale und arbeitsformspezifische Unterschiede in Bezug auf die inhaltlichen, formalen und sprachlichen Qualitätsaspekte.
1c	Bei den einzelnen Qualitätsaspekten der Maturaarbeit gibt es keine signifikanten Unterschiede aufgrund der disziplinären Zuordnung.
1d	Interdisziplinär ausgerichtete Arbeiten weisen keine signifikanten qualitativen Unterschiede zu Arbeiten mit klarer disziplinärer Zuordnung auf.

2.5.2 Institutionelle Ebene

Es ist davon auszugehen, dass die institutionellen Rahmenbedingungen, mit denen die Schülerinnen und Schüler während des Verfassens der Maturaarbeit konfrontiert sind, einen Einfluss auf die Qualität der Maturaarbeiten haben. Schon die Studie von Zillig (2004) zeigte, dass die Rahmenbedingungen zwischen den einzelnen Schulen differieren. So existieren je nach Schule beispielsweise unterschiedliche Vorgaben in Bezug auf die Wahl eines Themas oder einer Fragestellung.

Themenwahl

In der vorliegenden Untersuchung wird angenommen, dass die Freiheit bei der Themenwahl einen Einfluss auf die inhaltliche Qualität der Maturaarbeiten hat: Ausgehend von der Annahme, dass hohe Freiheit bei der Themenwahl zu höherem Interesse und zu höherer Motivation auf Seiten der Schülerin oder des Schülers führt, wird die Hypothese aufgestellt, dass sich grosse Freiheit bei der Themenwahl positiv auf die inhaltliche Qualität auswirkt (→ Hypothese 2a).

Tab. VI.2: Hypothesen zur Themenwahl

2a	Themenwahl
	Je höher die Freiheit bei der Themenwahl, desto besser die inhaltliche Qualität der Maturaarbeit.

Vorbereitung

Auf der Ebene der institutionellen Rahmenbedingungen spielen auch vorbereitende Angebote eine Rolle. Hierunter werden Angebote wie etwa Projekt- oder Intensivwochen zur Maturaarbeit, Kurse zum wissenschaftlichen Schreiben, Lernen am Projekt oder vorgängige Semesterarbeiten verstanden. Es ist davon auszugehen, dass solche Angebote sich positiv auf die Qualität der Maturaarbeiten, insbesondere auf die formale Qualität der Maturaarbeiten auswirken (→ Hypothesen 2b1 und 2b2). Da die Qualität solcher vorbereitender Veranstaltungen zwischen einzelnen Schulen oder auch je nach Lehrperson(en) beträchtlich divergieren kann, ist zu prüfen, ob es einen positiven Zusammenhang zwischen der wahrgenommenen Nützlichkeit dieser Veranstaltungen und der Qualität der Maturaarbeiten gibt (→ Hypothese 2b3).

Tab. VI.3: Hypothesen zur Vorbereitung

2b	Vorbereitung
2b1	Intensive Vorbereitungsveranstaltungen (Intensiv-/Projektwoche, LAP, Einführungskurse, ev. andere) wirken sich positiv auf die Qualität der Arbeit aus.
2b2	Die Maturaarbeiten derjenigen Schülerinnen und Schüler, welche lediglich Informationsveranstaltungen oder gar keine vorbereitende Veranstaltungen besucht haben, weisen tendenziell eine geringere Qualität auf.
2b3	Als nützlich wahrgenommene Vorbereitungsveranstaltungen wirken sich positiv auf die Qualität aus.

Realisierungsbedingungen

In der Studie von Zillig (2004) zeigte sich, dass die Zeitspanne, in welcher die Maturaarbeit verfasst werden muss, um bis zu zehn Monate variiert. In den Studien von Binder und Feller-Länzlinger (2003; 2004) äusserten sich die meisten der befragten Maturandinnen und Maturanden dahingehend, dass eine Bearbeitungszeitspanne von sechs Monaten ideal sei. Ebenso zeigte die Umfrage von Schmid (2002), dass die Schülerinnen und Schüler eine Zeitspanne von einem Jahr als zu lang und eine Zeitspanne von weniger als acht Monaten als ausreichend erachten. Ausgehend von diesen Ergebnissen wird die Hypothese aufgestellt, dass es – in Bezug auf die Qualität der Maturaarbeiten – eine ideale Bearbeitungszeitspanne gibt (→ Hypothese 2c1).

Verschiedene Untersuchungen zeigten, dass viele Maturandinnen und Maturanden die an ihrer Schule verfügbaren Wegleitungen, Richtlinien und Weisungen nicht als zufrieden stellend erachten (vgl. Davaud & Hexel, 2003; Pagnossin et al., 2005; Stocker, 2006). Davaud und Hexel (2003) stellten zudem fest, dass Schülerinnen und Schüler, die mit den Realisierungsbedingungen eher unzufrieden sind, auch den Nutzen der Maturaarbeit eher in Frage stellen. Vor diesem Hintergrund soll untersucht werden, ob transparentere Vorgaben, insbesondere auch transparente Bewertungskriterien zu höherer Qualität der Maturaarbeit führen (→ Hypothese 2c2, 2c3, 2c4). Neben der Transparenz der Vorgaben und Bewertungskriterien soll auch untersucht werden, ob die Zufriedenheit mit diesen einen Einfluss auf die Qualität der Maturaarbeit hat (→ Hypothese 2c5).

Tab. VI.4: Hypothesen zu den Realisierungsbedingungen

2c	Realisierungsbedingungen
2c1	Es gibt einen Zusammenhang zwischen der institutionell vorgegebenen Bearbeitungszeitspanne und der Qualität der Maturaarbeit.
2c2	Die Verfügbarkeit und Kenntnis eines Leitfadens führt zu einer höheren Qualität der Maturaarbeit.
2c3	Die wahrgenommene Transparenz der Abläufe wirkt sich positiv auf die Qualität der Arbeit aus.
2c4	Die wahrgenommene Transparenz der Bewertungskriterien wirkt sich positiv auf die Qualität der Arbeit aus.
2c5	Es gibt keinen Zusammenhang zwischen der Beurteilung der Bewertungskriterien und der Qualität.

2.5.3 Betreuungsebene

Maturandinnen und Maturanden werden beim Verfassen der Maturaarbeit in der Regel durch eine oder mehrere betreuende Lehrpersonen unterstützt. Im Prozess des Erstellens einer Maturaarbeit wirkt die betreuende Lehr-

person direkt oder indirekt auf die Qualität der Arbeit ein. Die Betreuungssituation hat indirekte Auswirkungen auf die Qualität der Maturaarbeiten, indem es die betreuende Lehrperson beispielsweise schafft, das Interesse und die Motivation der Schülerinnen und Schülern zu wecken. Es sind jedoch auch unmittelbare Einflüsse der Betreuungsebene auf die Arbeitsqualität denkbar, etwa wenn in Betreuungsprozessen direkt in die Arbeit eingegriffen wird. So wird unter anderem zu untersuchen sein, ob Maturaarbeiten von Maturandinnen und Maturanden, welche mit ihrer Betreuungssituation zufrieden sind, eine höhere Qualität aufweisen als Maturaarbeiten von unzufriedenen Schülerinnen und Schülern (→ Hypothese 3a). Es soll im vorliegenden Projekt insbesondere auch untersucht werden, inwiefern sich die Zufriedenheit mit der Unterstützung in einzelnen Bereichen auf die verschiedenen Qualitätsaspekte auswirkt (→ Hypothesen 3b, 3c).

Ausserdem wird angenommen, dass Schülerinnen und Schüler, welche ihre Betreuungsperson selbst auswählen können, eine Betreuungsperson wählen, die sie als kompetent und unterstützend einschätzen und dementsprechend mit der Qualität der Betreuung zufrieden sind (→ Hypothese 3d).

Im Gegensatz zu der Annahme, dass sich die Zufriedenheit mit der Betreuung positiv auf die Qualität der Maturaarbeiten auswirkt, geht die vorliegende Untersuchung davon aus, dass die Häufigkeit der stattfindenden Betreuungskontakte sowie die Form der Betreuung (persönlicher Kontakt, via E-Mail, individuelle Betreuung vs. Betreuung in Gruppen) keine Auswirkungen auf die Qualität der Maturaarbeit hat (→ Hypothese 3e).

Tab. VI.5: Hypothesen zur Betreuung

3	Betreuung
3a	Je besser die wahrgenommene Betreuung, desto besser die Qualität der Maturaarbeit.
3b	Die Zufriedenheit mit der Unterstützung bei der Formulierung der Fragestellung und der Unterstützung in Bezug auf den Inhalt der Arbeit korreliert positiv mit der inhaltlichen Qualität der Arbeit.
3c	Die Zufriedenheit mit der Unterstützung in Bezug auf formale Grundlagen sowie den Aufbau und die Gliederung der Arbeit korreliert positiv mit der formalen Qualität der Arbeit.
3d	Je höher die Freiheit bei der Betreuerwahl, desto positiver wird die Betreuung wahrgenommen.
3e	Zwischen der Häufigkeit sowie der Form der Betreuungskontakte und der Qualität der Maturaarbeit gibt es keinen Zusammenhang.

2.5.4 Individuelle Ebene

Die relevanten Faktoren auf der individuellen Ebene wirken sich stets direkt auf die Qualität der Maturaarbeit aus. Es darf jedoch nicht ausser Acht gelassen werden, dass die Faktoren auf der individuellen Ebene sehr oft durch die Umwelt und damit durch die institutionelle sowie insbesondere auch durch die Betreuungsebene beeinflusst werden.

Motivation und Selbstwahrnehmung

So gilt es zu untersuchen, ob sich die in die Maturaarbeit investierte Zeit – welche unter Umständen abhängig ist von der durch die Schule zur Verfügung gestellten Zeit sowie der zur Erarbeitung festgelegten Zeitspanne – auf die Qualität der Arbeiten auswirkt (→ Hypothese 4a1).

In Bezug auf die individuelle Ebene muss auch die Unterstützung durch das persönliche Umfeld der Maturandinnen und Maturanden thematisiert werden. In den Studien von Pagnossin et al. (2005) und Binder und Feller-Länzlinger (2003, 2004) zeigte sich, dass das Umfeld oder Drittpersonen bei der Themenwahl eine Rolle spielen

können; genauso ist auch vorstellbar, dass Maturandinnen und Maturanden beim Verfassen der Maturaarbeit durch das Umfeld oder Drittpersonen unterstützt werden. Es ist davon auszugehen, dass Maturaarbeiten von Maturandinnen und Maturanden, die neben der Unterstützung durch ihre Betreuungsperson, auch von anderer Seite her Unterstützung erfahren, tendenziell eine höhere Qualität aufweisen (→ Hypothese 4a2).

Ausserdem wird vermutet, dass Maturandinnen und Maturanden, die über Vorwissen in Bezug auf die Thematik ihrer Arbeit verfügen, Maturaarbeiten mit inhaltlich höherer Qualität verfassen als Maturandinnen und Maturanden, die sich an eine für sie eher neue Thematik wagen (→ Hypothese 4a3). Darüber hinaus ist zu prüfen, ob Schülerinnen und Schüler, die sich in Bezug auf ihre Kompetenzen als gut gerüstet wahrnehmen, tendenziell bessere Arbeiten eingereicht haben (→ Hypothese 4a4). Zudem ist zu klären, ob das erklärte Interesse der Schülerinnen und Schüler einen gerichteten Zusammenhang mit der Qualität der Maturaarbeit aufweist (→ Hypothese 4a5).

Tab. VI.6: Hypothesen zur Motivation und Selbstwahrnehmung

4a	Motivation und Selbstwahrnehmung
4a1	Der wahrgenommene Zeitaufwand korreliert positiv mit der Qualität der Arbeiten.
4a2	Unterstützung durch das eigene Netzwerk (Eltern, Geschwister und Peers) wirkt sich positiv auf die Qualität der Maturaarbeit aus.
4a3	Vorkenntnisse zum Thema wirken sich positiv auf die inhaltliche Qualität aus.
4a4	Die Kompetenzwahrnehmung im Hinblick auf Vorkenntnisse und Selbstständigkeit korreliert positiv mit der Qualität.
4a5	Das retrospektiv deklarierte Interesse korreliert positiv mit der Qualität.

Intrinsischer und extrinsischer Wert der Arbeit

Weiterhin gilt es zu untersuchen, ob die Qualität der Arbeit mit der Einschätzung des Nutzens sowie des Kompetenzzuwachses für die weitere Qualifikation korreliert. Dabei wird davon ausgegangen, dass eine hohe Qualität der Maturaarbeit auch auf Seiten des Schülers zu einer positiven Einschätzung des Nutzens der Arbeit sowie des Kompetenzzuwachses durch das Erstellen der Arbeit führt (→ Hypothese 4b1, 4b2).

Da den Schülerinnen und Schülern zum Zeitpunkt der Datenerhebung bereits eine Einschätzung der Qualität der Maturaarbeit in Form der Beurteilung durch die Lehrperson vorlag, muss davon ausgegangen werden, dass die Einschätzung des eigenen Kompetenzzuwachses in Abhängigkeit davon erfolgt (→ Hypothese 4b3).

Tab. VI.7: Hypothesen zum intrinsischen und extrinsischen Wert der Arbeit

4b	Intrinsischer und extrinsischer Wert der Arbeit
4b1	Die Qualität der Arbeit korreliert positiv mit der Einschätzung des Nutzens für die weitere Qualifikation.
4b2	Die Einschätzung des Nutzens korreliert positiv mit der Einschätzung des eigenen Kompetenzzuwachses.
4b3	Die Bewertung der Arbeit durch die Lehrperson korreliert positiv mit der Einschätzung des eigenen Lernzuwachses.

2.6 Untersuchungspopulation und Stichprobenziehung

2.6.1 Untersuchungspopulation

Die Grundgesamtheit der Untersuchung setzt sich aus allen Gymnasiastinnen und Gymnasiasten zusammen, die im Schuljahr 2006/07 ihre Maturaarbeit verfasst haben. Da die Dauer der gymnasialen Ausbildung sowie die Terminierung der Maturaarbeit variieren[41] (teilweise sogar innerschulisch), wurde zur Bestimmung der Stichprobe auf die Daten aller Gymnasiastinnen und Gymnasiasten, die sich im Schuljahr 2004/05 im 1. Bildungsjahr (10. Schuljahr) befanden, zurückgegriffen. Dies geschah unter der Annahme, dass diese Schülerinnen und Schüler im Schuljahr 2006/07 ihre Gymnasialausbildung abgeschlossen haben (vgl. Tabelle VI.8). Es kann aufgrund der vorliegenden Statistiken davon ausgegangen werden, dass die Grundgesamtheit der Untersuchung rund 21'980 Schülerinnen und Schüler in rund 1'044 Klassen umfasst. Im gesamtschweizerischen Durchschnitt umfasst die Klassengrösse also rund 21 Schülerinnen und Schüler pro Klasse (D-CH: 20,6; F-CH: 22,1; I-CH: 19,4).

Tab. VI.8: Grundgesamtheit der Untersuchung (Gymnasiastinnen und Gymnasiasten des 1. gymnasialen Bildungsjahres im Schuljahr 2004/05)

	Schweiz	D-CH	F-CH	I-CH
Frauen	12 357 (56,2 %)	6 706 (56,5 %)	4 834 (56,4 %)	817 (53,4 %)
Männer	9 623 (43,8 %)	5 166 (43,5 %)	3 743 (43,6 %)	714 (46,6 %)
Total	**21 980 (100 %)**	**11 872 (54 %)**	**8 577 (39 %)**	**1 531 (7 %)**

(Quelle: BfS)

2.6.2 Stichprobenziehung

Die Stichprobenziehung lehnt sich vom Ansatz her an das im Rahmen von EVAMAR I gewählte Verfahren an (vgl. Ramseier, 2005). Die Stichproben der Maturaarbeiten sowie der Schülerinnen- und Schülerbefragung basieren nicht auf einer Zufallsauswahl, sondern auf einer systematischen Ziehung. Die Ziehung der Stichprobe erfolgte in einem ersten Schritt auf der Ebene der Schulen. Als Grundlage wurde das *Verzeichnis der Schulen, deren Maturitätsausweise vom Bund und der Schweizerischen Konferenz der kantonalen Erziehungsdirektoren (EDK) anerkannt sind* (Stand: 1. März 2005) verwendet, das von der Schweizerischen Maturitätskommission geführt wird. Dieses Verzeichnis ist nach Kantonen sortiert und innerhalb der Kantone werden die Schulen alphabetisch nach dem (Haupt-) Standort aufgeführt. Die Population der Schulen umfasst insgesamt 169 Einheiten, wobei für die vorliegende Untersuchung alle Maturitätsschulen für Erwachsene sowie Schweizer Schulen im Ausland ausgeschlossen wurden, so dass die Population der für diese Untersuchung in Frage kommenden Schulen noch 158 Einheiten umfasst.

Um sicherzustellen, dass alle drei Sprachregionen angemessen vertreten sind, wurde eine Schichtung der Stichprobe nach Sprachregionen vorgenommen. Die Aufteilung der Stichprobe nach Sprachregionen bietet sich einerseits aus forschungslogischen Gründen an. Andererseits aber auch, da EVAMAR I zeigte, dass die Rahmenbedingungen bei der Umsetzung der Maturaarbeit in den drei Sprachregionen unterschiedlich ausgestaltet sind, dies insbesondere hinsichtlich der Wahlfreiheit in Bezug auf das Thema der Maturaarbeit. Auf die Bildung weite-

[41] An den meisten Schulen wird die Maturaarbeit im zweitletzten oder letzten Jahr erstellt (Zillig, 2004, S. 18).

rer Strata (z.B. Kantone, Schulgrösse usw.) wurde verzichtet, da davon ausgegangen wurde, dass Faktoren wie die Kantonszugehörigkeit oder Schulgrösse keinen Einfluss auf die Qualität der Maturaarbeiten haben.

Innerhalb der expliziten Strata wurden die Schulen so sortiert, dass Schulen aus gleichen Kantonen benachbart sind, woraus sich eine Bildung impliziter Strata ergibt. Die Ziehung erfolgte ausgehend von einer Zufallszahl. Es wurde jede x-te Schule gezogen, wobei x so gewählt wurde, dass die gewünschte Anzahl M der insgesamt N Schulen des jeweiligen Stratums gewählt wurde. Damit wird eine bessere Schätzung von Mittelwerten erreicht als mit einem reinen Zufallsverfahren (Lohr, 1999 zit. nach Ramseier, 2005, S. 6). Bei dieser Form der Ziehung beträgt die Selektionswahrscheinlichkeit für eine Schule M/N.

Tab. VI.9: Grundgesamtheit und Stichprobe der Gymnasien

Gymnasien	Grundgesamtheit	Stichprobe
Gesamtschweiz	158 (100 %)	30 (100 %)
D-CH	118 (74,4 %)	16 (53,3 %)
F-CH	35 (22,2 %)	12 (40 %)
I-CH	5 (3,2 %)	2 (6,7 %)

Nach der Ziehung der Schulen wurde pro Schule eine Stichprobe von Klassen gezogen, die im Schuljahr 2006/07 ihre Maturaarbeit verfassten. Es galt die Regel, dass bei Schulen mit bis zu zwei in Frage kommenden Klassen jeweils eine Vollerhebung durchgeführt wurde. Bei Schulen mit mehr als zwei Klassen wurden mittels Zufallsverfahren[42] zwei Klassen gezogen. Für die Schülerinnen- und Schülerbefragung wurden die ausgewählten Klassen komplett in die Stichprobe aufgenommen.

Die Ausführungen machen deutlich, dass die Wahrscheinlichkeit, als Maturandin oder Maturand für die Stichprobe gezogen zu werden, je nach Grösse der Schule variiert. Es kann also nicht von einer repräsentativen Stichprobe gesprochen werden, da Schülerinnen und Schüler aus kleineren Schulen überproportional vertreten sind. Ausschlaggebender Grund für das gewählte Verfahren der Stichprobenziehung war es jedoch, durchgängig Fallzahlen zu erreichen, die es erlauben, einen Bezug zwischen der Wahrnehmung der Schülerinnen und Schüler und den formalen schulischen Rahmenbedingungen herzustellen. Gewichtungen wurden im Rahmen der Analyse lediglich dann vorgenommen, wenn ein expliziter Vergleich zu Daten aus dem Projekt EVAMAR I stattfindet und dabei der Aspekt der soziodemographischen Vergleichbarkeit berücksichtigt werden muss.

Die Untersuchung von Pagnossin et al. (2005) zeigt, dass rund drei Viertel aller Maturaarbeiten theoretischer oder empirischer Natur sind. Die anderen Arbeitstypen sind wesentlich seltener. Auf der Grundlage dieser Information und der Überlegung, dass ein Einbezug von ausserschulischen Aktivitäten, soziokulturellen Projekten sowie künstlerischen, audiovisuellen und materiellen Produktionen zu Problemen der Vergleichbarkeit zwischen den Arbeiten führen würde, wurde die Ziehung der Maturaarbeiten auf theoretische und empirischen Arbeiten beschränkt.

Um vor der Ziehung des Samples für die Evaluation der Maturaarbeiten zu wissen, welche Arbeiten in Frage kommen, wurden die Schulen aufgefordert, die Maturaarbeiten der gewählten Klassen danach zu kategorisieren, ob sie theoretischer, empirischer oder anderer Natur seien. In der Praxis der Durchführung der Untersuchung zeigte sich alsbald, dass es nicht möglich ist, Maturaarbeiten trennscharf zu kategorisieren. In unsere Stichprobe gelangten immer wieder auch Arbeiten, die Berichte zu technischen oder kreativen Produktionen oder durchge-

[42] Es wurden zwei Zufallszahlen y1 und y2 zwischen 1 und x (wobei x = Anzahl Klassen einer Schule) generiert und die Klassen an y1. und y2. Stelle innerhalb der von der Schule eingereichten Klassenliste gezogen.

führten Veranstaltungen umfassten, weil diese von den angefragten Schulen sehr oft als empirische Arbeiten klassifiziert wurden.

Auf der Basis dieser Liste sowie aufgrund des Rücklaufs der Fragebogen der Schülerinnen- und Schülerbefragung wurden schliesslich in einem letzten Schritt aus jeder der im Sample vertretenen Klassen mittels Zufallsverfahren eine bestimmte Anzahl Schülerinnen und Schüler gezogen, von denen uns sowohl ein Fragebogen als auch eine Maturaarbeit vorlagen.

Dieses Verfahren der systematischen Stichprobenziehung geht mit einem Effizienzverlust einher. Der Designeffekt misst diesen Genauigkeits- oder Effizienzverlust einer komplexen Stichprobe – wie der hier generierten geschichteten Klumpenstichprobe – im Vergleich zu einer einfachen Zufallsstichprobe und gibt an, um wie viel sich die effektive Stichprobengrösse durch das Design der Stichprobe verringert. Der Designeffekt ist von der Homogenität (Intraklassenkorrelation) und der Grösse der einzelnen Klumpen (Schulen, Klassen) abhängig: Schülerinnen und Schüler innerhalb einer Schule oder Klasse beeinflussen sich gegenseitig hinsichtlich der Qualität der Maturaarbeit und treffen damit homogenere Aussagen als Personen innerhalb einer zufällig zusammengestellten Gruppe. Will man zuverlässige Aussagen mit den erhobenen Daten treffen, muss die Stichprobe dementsprechend vergrössert werden, um trotz Designeffekt eine repräsentative Heterogenität zu erfassen.

Für die geplante Untersuchung gehen wir von einem Designeffekt von deff=4,5 aus. Zum Vergleich: Die Civic-Education-Study ging von einem deff=6,6 aus, PISA 2000 Mathematik von einem deff=5,5. Im Gegensatz zu unserer Untersuchung zur Qualität der Maturaarbeit spielt bei diesen beiden Studien der Klasseneffekt eine viel grössere Rolle als beim Verfassen einer Maturaarbeit. Da das Verfassen der Maturaarbeit organisatorisch nicht an den Kontext einer bestimmten Schulklasse gebunden ist, die als Lerngruppe mit einzelnen Lehrpersonen interagiert, gehen wir in diesem Fall von einem geringeren Designeffekt aus.

Zur Bestimmung der Stichprobengrösse legten wir fest, dass der Mittelwert der Stichprobe im Bereich von ±0,2 Standardabweichung (SD) des Mittelwertes der Grundgesamtheit liegen soll, um noch mit einiger Genauigkeit von der Stichprobe auf die Grundgesamtheit schliessen zu können. Um eine ausreichende Genauigkeit (95%) von Aussagen über Mittelwerte zu erhalten, muss also eine Stichprobengrösse von ca. 100 zugrunde gelegt werden.[43] Berücksichtigt man den angenommenen Designeffekt von deff=4.5, ergibt sich eine Stichprobengrösse von rund n=450.

2.6.3 Datenerhebung

Nachdem in einem ersten Schritt die Stichprobe auf der Ebene der Schulen gezogen wurde, wurden diese Anfang 2007 angeschrieben und gebeten, an der Untersuchung teilzunehmen. Von den n=30 Schulen (D-CH: 16, F-CH: 12, I-CH: 2) mussten 13 Schulen (D-CH: 7, F-CH: 6) ersetzt werden, weil sie entweder keine gymnasiale Ausbildung mehr anbieten (2 Schulen) oder weil sie explizit keine Teilnahme am Projekt wünschten (11 Schulen).

In der Zeit von April bis Juni 2007 wurde der Schülerinnen- und Schülerfragebogen an die Schulen versandt. Von den insgesamt 1'073 Schülerinnen und Schüler, die so kontaktiert werden konnten, beantworteten n=920 Schülerinnen und Schüler (w: 542; m: 378) den Fragebogen, das entspricht einer Rücklaufquote von 85,74%. Das Altersspektrum der Befragten erstreckte sich von 17 bis 25 Jahren; wobei das Durchschnittsalter zum Zeitpunkt der Erhebung 18,9 Jahre (SD=1.01) betrug.

Die Rücklaufquote ist nicht in allen drei Sprachregionen gleich hoch. In der französischsprachigen Schweiz konnte der Fragebogen nur bei acht Schulen im Klassenverband ausgefüllt werden, da im letzten Trimester vor der Maturaprüfung der Grossteil des normalen Unterrichts durch persönliche Vorbereitungszeit substituiert wird. Dasselbe trifft für eine der Tessiner Schulen zu. An diesen Schulen wurden, wenn nicht anders möglich,

[43] Die unter den beschriebenen Bedingungen notwendige Stichprobengrösse berechnet sich nach der Formel $n=(SD/SE)^2$, wobei SD die Standardabweichung der Grundgesamtheit ist und SE der Standardstichprobenfehler. Wenn man eine 95%ige Genauigkeit annimmt, ergibt sch aus den genannten Bedingungen, dass 1.96 SE=0,2 SD. Damit ergibt sich für n ein Wert von 96.

die Schülerinnen und Schüler von uns direkt angeschrieben. Es zeigte sich relativ schnell, dass der Rücklauf der Fragebogen an diesen Schulen geringer war. Vor dem Hintergrund dieser Information ist ebenfalls die Verteilung der befragten Personen auf die Sprachregionen zu betrachten. Es entfallen 59,3% der Befragten auf Maturitätsschulen in der Deutschschweiz, 34,7% besuchen Schulen in der Romandie und 6% im Kanton Tessin. Gegenüber der Grundgesamtheit, in der nur 54% Prozent der Maturandinnen und Maturanden aus der Deutschschweiz stammen, dagegen 39% aus der Romandie und 7% aus dem Tessin, ist die Stichprobe also leicht verzerrt, die Aussagekraft der regionalen Resultate wird dadurch teilweise relativiert.

Ausgehend von den n=920 Schülerinnen, deren Fragebögen uns vorliegen, wurden mittels des oben beschriebenen Verfahrens n=437 Schülerinnen und Schüler ausgewählt, die eine theoretische oder empirische Maturaarbeit verfasst haben.[44] Diese Maturaarbeiten wurden von Juni bis Dezember 2007 einer Evaluation mittels des Kriterienrasters unterzogen. Hierbei waren insgesamt 42 Rater beteiligt.

3 Institutionelle Rahmenbedingungen

Im folgenden Teil werden die Rahmenbedingungen für die Erstellung der Maturaarbeiten vergleichend dargestellt. Die Ziel- und Zweckbestimmungen, die Vorgaben zur Arbeitsorganisation, zur Betreuung sowie zur Bewertung finden sich je nach Kanton und je nach Schule auf unterschiedlichen Ebenen. Diese Uneinheitlichkeit entspricht der Logik eines föderalistischen Systems, zu dessen Eckpfeilern das Recht der kantonalen Selbstgesetzgebung gehört. Unumschränkt ist diese Autonomie indes nicht, sie kann durch die eidgenössische Gesetzgebung eingeschränkt und unter Anpassungsdruck gesetzt werden. Die Analyse der Rahmenbedingungen geht deshalb von der Eidgenossenschaft als Makroebene aus, um sich dann von der kantonalen Mesoebene her der Mikroebene der im Untersuchungssample vertretenen Einzelschulen zuzuwenden.[45]

3.1 Übergeordnete Vorgaben

3.1.1 Zielsetzungen auf nationaler Ebene

Auf nationaler Ebene legt das MAR 95 Rahmenbedingungen zur Maturaarbeit fest. In Artikel 10 werden die Vorgaben für die Maturaarbeit wie folgt definiert:

> Art. 10 Maturaarbeit
>
> Schülerinnen und Schüler müssen allein oder in einer Gruppe eine grössere eigenständige schriftliche oder schriftlich kommentierte Arbeit erstellen und mündlich präsentieren.

Die Vorgaben betreffen also vordergründig nur arbeitsorganisatorische und instrumentelle Komponenten. Es fehlen explizite Hinweise sowohl auf die Zielsetzungen und den Stellenwert der Maturaarbeit als auch auf den wissenschaftspropädeutischen Bezug und mögliche Inhalte. Ein zentrales Element von Artikel 10 stellt die *Eigenständigkeit* dar. Das Konzept der Eigenständigkeit weist einerseits auf Arbeitsformen hin, welche nicht den üblichen gymnasialen Unterrichtsmustern entsprechen und es weist andererseits auf die Autonomie und Kreativität hin, welche von den Schülerinnen und Schülern im Rahmen der Maturaarbeit erwartet wird.

Weitere Erwähnung findet die Maturaarbeit in den Artikeln 15 und 20 des MAR 95: Zum einen findet sich hier die Bestimmung, dass bei der Bewertung der Maturaarbeit die erbrachten schriftlichen wie auch die mündlichen Leistungen zu berücksichtigen sind. Zum anderen lautet die Vorgabe, dass das Thema und die Bewertung der

[44] Die tatsächliche Anzahl untersuchter Maturaarbeiten beträgt n=406. Diese Zahl ist etwas geringer als die Anzahl der untersuchten Schülerinnen und Schüler. Dies ist darauf zurückzuführen, dass einige Schülerinnen und Schüler der Stichprobe ihre Maturaarbeit gemeinsam verfassten.

[45] Vorgaben und Reglemente, welche nach 2006 in Kraft getreten sind, werden nicht berücksichtigt, da sich das Sample auf die Vorgaben von 2006 bezieht.

Maturaarbeit im Maturitätsausweis anzuführen sind. Allerdings war diese Bewertung bisher für den gymnasialen Schulabschluss rechnerisch nicht bedeutsam und insofern hat die Angabe von Thema und Note lediglich eine symbolische Funktion erfüllt. Weitere Hinweise auf Anforderungen, die durch das Erarbeiten einer Maturaarbeit erreicht werden sollen, werden in den allgemeinen Zielbestimmungen des MAR 95 (Art. 5) gemacht – ohne dass jedoch direkt auf die Maturaarbeit Bezug genommen wird. In Ansätzen werden Ansprüche an eine Wissenschaftspropädeutik innerhalb des gymnasialen Lernens geltend gemacht, indem als Ziel die „Einsicht in die Methodik wissenschaftlicher Arbeit" (Art. 5, Abs. 2) festgehalten wird. In Bezug auf die Zeitanteile heisst es in Art. 11: „Für den Wahlbereich Schwerpunkt- und Ergänzungsfach sowie Maturaarbeit sind 15-25% Zeitanteile vorzusehen."

Als weitere Quelle auf nationaler Ebene ist der Rahmenlehrplan für Maturitätsschulen (EDK, 1994) zu nennen. Dieser enthält keine eigentlichen Inhaltsvorschriften, sondern lediglich übergeordnete Zielsetzungen, welche für alle Maturandinnen und Maturanden generelle Gültigkeit haben. So wird beispielsweise unter der Rubrik „allgemeine Ziele der Maturitätsbildung" festgehalten, dass Gymnasiastinnen und Gymnasiasten „auf der instrumentellen Ebene lernen [...] Theorien zu formulieren, Hypothesen aufzustellen, sie zu entkräften oder zu verifizieren und daraus Schlüsse zu ziehen" (EDK, 1994, S. 10f.). Die Schülerinnen und Schüler sollen weiterhin mit den methodischen Grundlagen wissenschaftlichen Arbeitens vertraut gemacht werden. Die Ausgestaltung und Konkretisierung der Inhalte der Lehrpläne bleibt aber in der Regel in der Hand der Schulen.

3.1.2 Zielsetzungen auf kantonaler Ebene

Kantonale Bestimmungen zur Funktion der Maturaarbeiten sind im Untersuchungssample[46] auf unterschiedlichen gesetzlichen Ebenen festgehalten. Die meisten Kantone übernehmen die wenig konkreten Vorgaben des MAR 95 in ihre kantonalen Gesetze oder Verordnungen; mitunter werden diese Verordnungen um weitere Einzelheiten ergänzt. Eine weitere Quelle auf dieser Mesoebene sind die kantonalen Lehrpläne, welche teilweise zusätzlich Präzisierungen zu den Zielsetzungen enthalten

Zielbestimmungen in kantonalen Rechtstexten

Einzelschulübergreifende Zielbestimmungen für die Maturaarbeiten, welche signifikant über die Forderungen des MAR 95 hinausgehen, kennt nur ein kleiner Teil der im Untersuchungssample vertretenen Deutschschweizer Kantone. So bestehen in vier Kantonen des Untersuchungssamples (nämlich: Basel-Landschaft, Basel-Stadt, Luzern, Obwalden) eigene Texte, die sich nebst einer Zieldefinition auch auf eine detaillierte Regelung der Form sowie der Abläufe und Resultate der Maturaarbeit beziehen. Die Inhalte dieser detaillierten Zielformulierungen zeigen eine grundlegende Übereinstimmung in Bezug auf das wissenschaftspropädeutische bzw. methodisch-systematische Vorgehen bei der Erstellung der Maturaarbeit sowie der Eigenständigkeit bei der Erarbeitung. Sowohl bezüglich der weiteren Inhalte und des Detaillierungsgrads der einzelnen Regelungen bestehen jedoch erhebliche Unterschiede von Kanton zu Kanton.

[46] Es sind dies die 18 Kantone, nämlich: Aargau, Basel-Landschaft, Basel-Stadt, Bern, Fribourg, Genf, Graubünden, Jura, Luzern, Neuenburg, Obwalden, St. Gallen, Tessin, Thurgau, Waadt, Wallis, Zug und Zürich.

Beispiel 1: Reglement für die Maturaarbeit an den Gymnasien im Kanton Basel-Stadt (vom 2. Februar 2001)

§ 1. Ziele

(1) Die Schülerin und der Schüler sollen für die Maturaarbeit anhand eines Themas eine komplexe Aufgabe formulieren, analysieren und dafür Lösungen erarbeiten. Sie oder er soll dabei angemessene Methoden anwenden und geeignete Hilfsmittel einsetzen.

(2) Die Schülerin und der Schüler sollen insbesondere:

- selbstständig ein Thema suchen, es eingrenzen, dazu relevante Fragestellungen entwickeln und es nach Absprache mit einer schulinternen Lehrkraft gemäss § 10 festlegen,
- selbstständig Informationen beschaffen,
- sich differenziert mit einer Materie auseinander setzen,
- das Thema formal, logisch aufgebaut und sprachlich korrekt präsentieren,
- ein Objekt gestalten und den Werdegang kognitiv nachvollziehen,
- die eigene Arbeit mündlich vorstellen und erläutern,
- mit Fachpersonen zielgerichtet zusammenarbeiten,
- je nach Thema interdisziplinäre Fragestellungen verfolgen,
- über längere Zeit mit Neugierde und persönlichem Interesse an einem Thema arbeiten.

In den meisten Kantonen beschränkt sich der Gesetzgeber auf die Wiederholung der Bestimmungen von Art. 10 des MAR 95 und erwähnt die jeweils geltenden Promotionsbestimmungen.

Beispiel 2: Graubünden. Verordnung über das Gymnasium im Kanton Graubünden (vom 6. Juli 1999)

Art.16 Maturaarbeit

(1) Schülerinnen und Schüler müssen alleine oder in einer Gruppe eine den Bestimmungen des Maturitätsanerkennungsreglementes entsprechende, eigenständige schriftliche oder schriftlich kommentierte Maturaarbeit erstellen und mündlich präsentieren.

(2) Bei der Maturaarbeit werden die erbrachten schriftlichen und mündlichen Leistungen beurteilt.

(3) Für das Maturitätszeugnis wird die in der Maturaarbeit erreichte Note den Prädikaten „hervorragend“, „sehr gut“, „gut“, „befriedigend“, „genügend“ oder „ungenügend“ zugewiesen.

Dies gilt auch für die Situation in der Romandie und im Tessin. In den Kantonen Fribourg, Genf, Jura, Neuenburg, Tessin, Waadt und Wallis bestehen zwar schulübergreifende Rechtstexte, auf welche sich die Einzelschulen in ihren Richtlinien mehr oder weniger direkt beziehen, die Zielbestimmungen gehen aber nur unmerklich über die Vorgaben des MAR hinaus.

Zusammenfassend kann gesagt werden, dass die Maturaarbeit in allen Kantonen gesetzgeberisch verankert ist, die Verankerung erfolgt jedoch auf unterschiedlichen gesetzlichen Ebenen. Dies kann dahin gehend ausgelegt werden, dass der Definition von Bildungszielen in den kantonalen Rechtsetzungen unterschiedliche Prioritäten zugemessen werden. Vom Parlament verabschiedete Gesetze symbolisieren eine höhere demokratische Legitimation, eine breitere Abstützung und einen formellen Vorrang gegenüber den Verordnungen oder Dekreten, die durch die Exekutive erlassen werden. Andererseits darf nicht ausser Acht gelassen werden, dass Verordnungen – aufgrund der kürzeren Entscheidungswege – anpassungsfähiger und flexibler sind. Ausserdem muss auch kritisch angemerkt werden, dass die juristische Verankerung von Ziel- und Zweckbestimmungen stets auch vom Rechtsetzungsprozess selbst gesteuert wird, wobei dieser Verankerung allerdings auch ein erhebliches Mass an politischer Kontingenz eigen ist.[47] Letztlich bleibt es in den meisten Fällen den Schulen selbst überlassen, die

[47] Für eine exakte Interpretation dieser Prozesse müssten die Begleitmaterialen des Rechtssetzungsprozesses genauer untersucht werden. Dies würde den Rahmen dieser Analyse jedoch sprengen.

Zielsetzungen für die Maturaarbeit genauer zu bestimmen und den Prozess ihrer Realisierung durch Reglemente und Wegleitungen zu konkretisieren.

Zielbestimmungen in kantonalen Lehrplänen

Schulübergreifende Vorgaben zur Maturaarbeit sind in den kantonalen Lehrplänen nur vereinzelt zu finden. So macht etwa der Kanton Aargau Angaben zu Organisation und Zielen der Maturaarbeit. Im Lehrplan des Kantons Bern finden sich Aussagen zum generellen Ziel von Maturaarbeiten, detaillierte Ausführungen werden jedoch den Gymnasien überlassen. Ebenfalls einleitende Bemerkungen zur Maturaarbeit finden sich in den kantonalen Lehrplänen der Kantone Basel-Stadt und Thurgau. In den meisten kantonalen Lehrplänen werden Maturaarbeiten nicht oder nur mit einem Verweis auf die Auslegung durch die einzelnen Gymnasien erwähnt.

3.2 Vorgaben auf der Ebene der Einzelschule

Die Analyse stützt sich auf die von den einzelnen Gymnasien mit einem Verbindlichkeitsanspruch formulierten und dokumentierten Informationen.[48] Dabei lag der Fokus auf der Schule als organisatorische Einheit – wohl wissend, dass im Fachunterricht oder in spezifischen Veranstaltungen oft weitere Erläuterungen und Fingerzeige geliefert werden, die im Rahmen einer nationalen Studie nicht berücksichtigt werden können. Die hier vorliegenden Dokumente stellen jedoch die wichtigsten Grundlagen für eine Qualitätsanalyse dar, weil sie den Anspruch verkörpern, den die jeweilige Schule als gymnasiale Institution an sich selbst stellt und mit dem sie sich nach Innen und nach Aussen präsentiert.

3.2.1 Zuständigkeiten und Richtlinien

Grundsätzlich ist festzuhalten, dass beinahe knapp die Hälfte aller untersuchten Schulen (43%) eigenständige Rechtsgrundlagen für die Maturaarbeit präsentieren. Hierbei handelt es sich meist um Reglemente, Promotionsordnungen oder verabschiedete Richtlinien. Die Genehmigungsinstanz für diese Rechtsvorgaben stellt in der Regel die Schulkommission des Gymnasiums dar. Im Unterschied zur Deutschschweiz sind schulspezifische Rechtsgrundlagen in der lateinischen Schweiz kaum verbreitet.

Elf der 30 untersuchten Gymnasien benennen spezifische Kommissionen oder Arbeitsgruppen, die für die Koordination und die Prozesse der Durchführung der Maturaarbeit schulweit verantwortlich sind. Diese Kommissionen bestehen zumeist aus Lehrpersonen und aus Mitgliedern der Schulleitung. Bei den übrigen Schulen sind keine solchen Gremien benannt. Bei sämtlichen Gymnasien der lateinischen Schweiz (mit Ausnahme der Kantonsschulen des Kantons Fribourg) liegt die Verantwortung für die Durchführung jeweils bei der zuständigen Schulleitung.

Ausnahmslos alle Gymnasien haben Leitfäden oder Richtlinien entworfen, die weitere Auskünfte über die Organisation, den Arbeitsprozess und die zu erwartenden Resultate geben sollen. Die Inhalte dieser Papiere variieren stark und reichen von rudimentären Kurzinformationen bezüglich der Leitidee, der Begleitung und Benotung bis hin zu einer Vielzahl von detaillierten Informationen, die sowohl übergreifende Themen ansprechen wie auch über spezifische Vorgaben je nach Fachbereich Auskunft geben. Die Mehrzahl der Gymnasien stellt mehr oder weniger ausführliche Unterlagen bereit, in denen die Bereiche *Arbeitsorganisation*, *Themenwahl*, *formale Vorgaben*, *Präsentation* und *Beurteilung* sowie die Termine behandelt werden.

Für die Kantonsschulen der Romandie und des Tessins gilt, dass auch auf die verschiedenen Fachbereiche bezogene Informationen vorliegen. Bedingt ist dies durch die Seminarorganisation, bei der die Oberthemen von Lehrpersonen in spezifischen Veranstaltungen mit einer beschränkten Anzahl Teilnehmenden vertieft werden. Auf-

[48] Da den beteiligten Gymnasien Anonymität zugesichert wurde, werden im Folgenden die Namen der Kantonsschulen nicht genannt, sondern es wird lediglich eine Einteilung nach Sprachregion vorgenommen.

grund der zentralen Rolle, welche die begleitende Lehrperson in dieser Organisationsform einnimmt, ist der Konkretisierungsgrad der schriftlich vorliegenden Information in diesen Gymnasien eher gering.

Artikel 10 des MAR 95 besagt, dass Maturaarbeiten allein oder in einer Gruppe erstellt werden können. Sieben Schulen beschränken die Grösse von Gruppen auf zwei oder drei Personen. In den anderen Schulen werden hierzu keine genaueren Angaben gemacht.

3.2.2 Gymnasiale Zielbestimmungen

Beinahe alle Gymnasien formulieren Ziele, die durch das Verfassen von Maturaarbeiten erreicht werden sollen. Gemäss diesen Zielformulierungen steht in der Regel das Einüben in wissenschaftliches Reflektieren und in die Methodik im Fokus der gymnasialen Maturaarbeit. Zudem wird in fast allen Zielsetzungen der Typus des „projektbezogenen Arbeitens“ genannt, bei dem die Selbstorganisation im Vordergrund steht. Bei der untenstehenden Auflistung fällt auf, dass die Mehrheit der untersuchten Deutschschweizer Schulen dem Aspekt der Interdisziplinarität keine Priorität beimisst, diese jedoch in beinahe allen Zielformulierungen der Schulen der französischsprachigen Schweiz explizit als Ziel formuliert ist. Auch findet sich der Aspekt der Wissensakkumulation ausschliesslich in den Zielformulierungen der Schulen der Romandie (vgl. Tabelle VI.10).

Tab. VI.10: Inhalte der Zielsetzungen, 2006 (N=30)

	Autonomie (Eigen-/ Selbstständig)	Kommunikation (schriftlich + mündlich) Präsentation	Methoden & wissens. Reflexion	Organisation Techniken und Termine	Kreativität / Originalität	Interdis-ziplinarität	Wissens-akkumulation
D CH 1	■	■	■				
D CH 2	■		■		■		
D CH 3	■	■	■	■	■		
D CH 4							
D CH 5	■	■	■	■			
D CH 6							
D CH 7	■		■	■			
D CH 8			■		■		
D CH 9	■			■			
D CH 10	■	■	■	■			
D CH 11							
D CH 12	■	■	■	■			
D CH 13	■	■	■	■			
D CH 14	■	■	■	■			
D CH 15	■	■	■	■	■		
D CH 16			■	■	■		
F CH 1		■	■	■	■		
F CH 2	■	■	■	■	■	■	
F CH 3		■	■				
F CH 4		■	■	■	■	■	■
F CH 5							

	Autonomie (Eigen-/ Selbstständig)	Kommunikation (schriftlich + mündlich) Präsentation	Methoden & wissens. Reflexion	Organisation Techniken und Termine	Kreativität / Originalität	Interdis-ziplinarität	Wissens-akkumlation
F CH 6	■	■	■	■	■		
F CH 7	■	■	■			■	■
F CH 8	■	■	■	■	■	■	■
F CH 9	■	■	■			■	■
F CH 10	■	■	■			■	■
F CH 11			■	■	■	■	
F CH 12	■	■	■	■	■	■	■
I CH 1		■	■			■	
I CH 2			■	■			
Total Nennungen	21	19	25	18	12	9	6

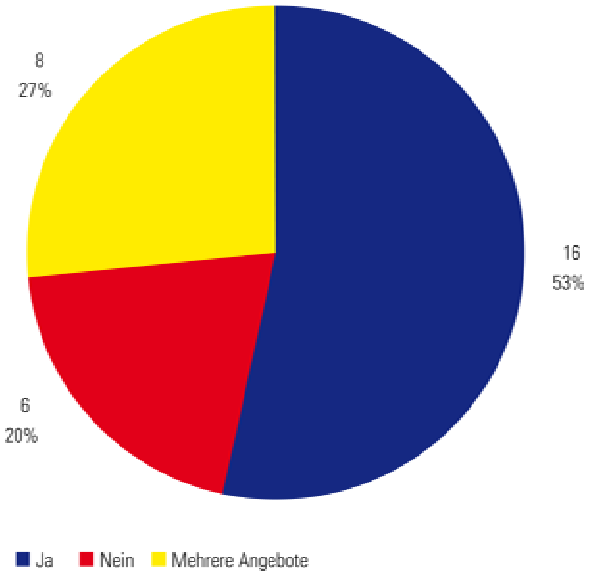
Hervorzuheben ist ausserdem, dass die im MAR 95 verlangte Eigenständigkeit in einigen Zielformulierungen die semantische Transformation hin zur *Selbstständigkeit* erfährt. Der auf ersten Blick unmerkliche Unterschied – beide Formulierungen zielen prinzipiell auf die Autonomie im Schaffensprozess – kann doch wichtige Differenzen beim Grad der kreativen Eigenleistungen der Schülerinnen und Schüler deutlich machen: Die eigenständige Reflexionsarbeit setzt selbstständige Informationssuche und -verarbeitung zwar voraus, ist aber nicht gleichbedeutend mit ihr.[49] Bemerkenswert ist weiterhin, dass in verschiedenen Kantonsschulen der Deutschschweiz (fünf von 16) explizit von Kreativität und Originalität in Bezug auf die Endprodukte die Rede ist. So wird etwa erwähnt, dass mit der Maturaarbeit „die Schaffung von Bedeutsamem" angestrebt wird oder dass die Maturaarbeit „originell und in wesentlichen Zügen neuartig sein" soll. Eine weitere Besonderheit stellt der Fokus auf ein Arbeitsprodukt dar. So sieht eine Kantonsschule vor, dass das Produkt der Maturaarbeit eine „von aussen wahrnehmbare Darstellung mit kommunikativem Wert in anspruchsvoller Form" sein soll.

3.2.3 Vorbereitung und Themenwahl

Zur Vorbereitung der Schülerinnen und Schüler auf die Maturaarbeit machen 24 von 30 Gymnasien spezifische Angebote (vgl. Abbildung VI.4). Dabei handelt es sich oft um klassenübergreifende Informationsveranstaltungen, in denen über die Ziele, Arbeitsprozesse und Termine informiert wird. Bei der weiteren Vorbereitung und in Bezug auf die Themenwahl sind jedoch grundlegende Unterschiede zwischen Schulen auszumachen. Acht der 30 in der Stichprobe vertretenen Schulen bieten eine erweiterte Angebotspalette: Beispielsweise können Schülerinnen und Schüler nebst der allgemeinen Information auch Kurse im Schreiben von Arbeiten besuchen. Weiter besteht in einem Fall die Möglichkeit, die Projektmethode durch Vorläuferarbeiten und im Rahmen von Unterrichtseinheiten einzuüben, und die Ergebnisse im Rahmen der Maturaarbeit auszubauen.

[49] Ähnliches festgestellt haben bereits Bonati und Hadorn (2007, S. 17).

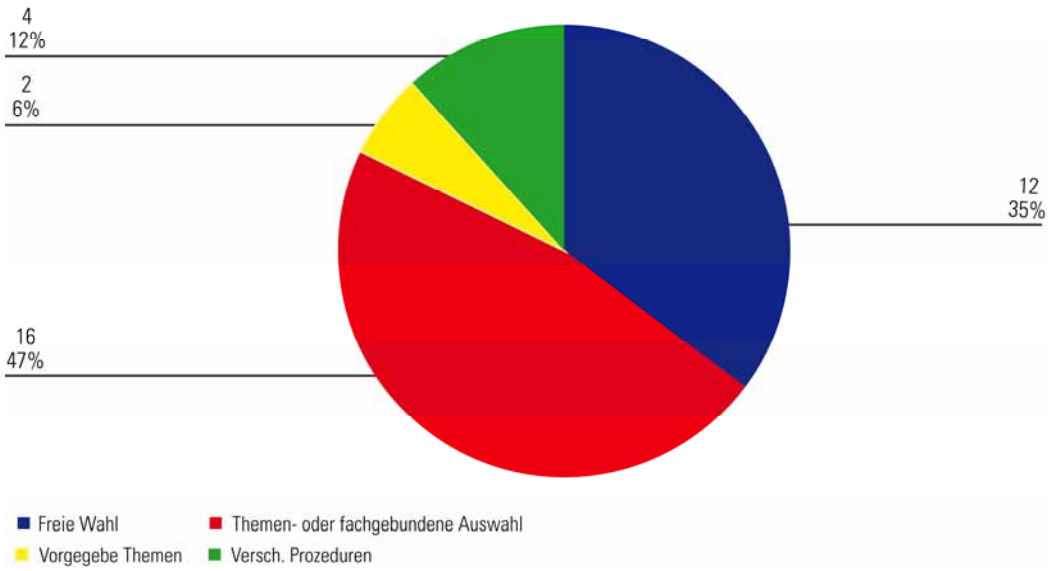
Abb. VI.4: Gymnasien und spezifische Vorbereitungsveranstaltungen



Die Kantonsschulen der lateinischen Schweiz organisieren die Vorbereitung vorwiegend in Seminaren in denen ein – meist von Lehrpersonen – vorgängig festgelegtes Thema behandelt wird. Innerhalb dieser Seminare, in denen sich Gruppen von bis zu 15 Schülerinnen und Schüler pro betreuende Lehrperson einschreiben können, wählen die Schülerinnen und Schüler dann ihre spezifischen Arbeitsthemen aus. Diese thematische Organisation des Arbeitsprozesses führt zu einer Wahlpflicht bezüglich möglicher Arbeitsthemen: Es wird der inhaltliche Rahmen vorgegeben, aus welchem die Schülerinnen und Schüler ihren zu bearbeitenden Gegenstand auswählen können.

Eine weitere Ausprägung in Bezug auf die Themenwahl besteht darin, dass Gymnasien ihren Schülerinnen und Schülern weitgehende Freiheiten lassen. Etwa ein Drittel der untersuchten Schulen verzichtet gar explizit auf eine schulfachbezogene Ausrichtung und damit unter Umständen auch auf eine Betreuung durch Lehrpersonen, welche über die fachlichen Kompetenzen verfügen. In knapp der Hälfte der Gymnasien besteht die Vorschrift, dass die Maturaarbeit in Verbindung mit einem oder mehreren Schulfächern stehen muss.

Abb. VI.5: Möglichkeiten der Themenwahl (N=30)



Die Möglichkeit, eine interdisziplinäre Thematik zu wählen, wird in keiner der im Sample enthaltenen Schulen ausgeschlossen. Explizit erwähnt wird Interdisziplinarität jedoch nur in den Zielvorgaben der Westschweizer Schulen (siehe Kap. 3.2.2). Eine Minderheit von Gymnasien erlaubt mehrere Verfahren – sowohl die freie bzw. frei-fachgebundene als auch die Themenwahl nach Vorgabe der Lehrpersonen. In allen untersuchten Gymnasien wird das gewählte Thema von den betreuenden Lehrpersonen bewilligt, bevor es von den Schülerinnen und Schüler bearbeitet werden kann.

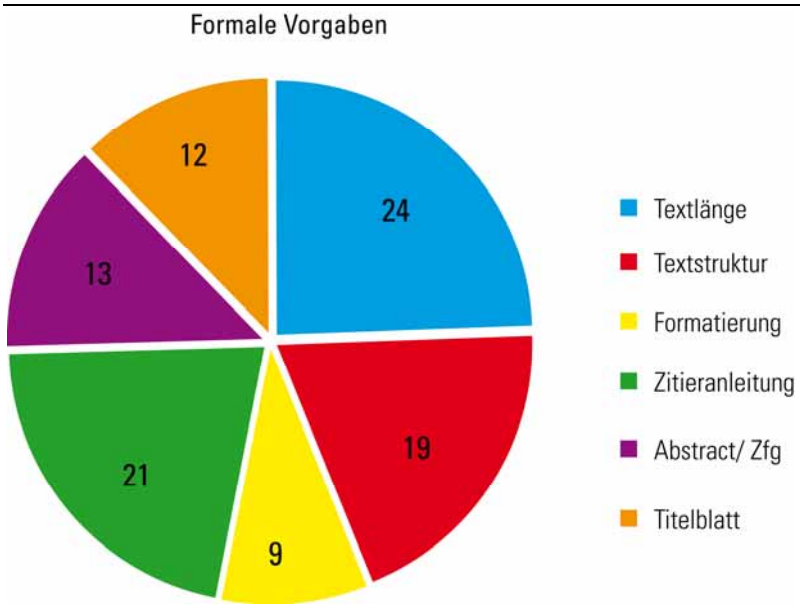
3.2.4 Formale Vorgaben

Die meisten Schulen verfügen über formale Richtlinien zur Gestaltung der Maturaarbeit. Diese Regelungen enthalten unterschiedliche Vorgaben bezüglich der Textlänge, des Aufbaus, des Layouts und der Gestaltung (vgl. Abbildung VI.6). Formale Vorgaben sind am häufigsten vorzufinden in Bezug auf die Textlänge. Hierzu machen 24 von 30 Gymnasien Angaben, wobei die Vorgaben mit Angaben zwischen 10-30 Seiten sehr stark variieren.[50] Drei der untersuchten Gymnasien (alle in der Deutschschweiz) verweisen zudem darauf, dass die minimale Textlänge mit den Vorgaben des Fachs, in dem die Arbeit geschrieben wird, abgestimmt werden muss. Sechs Schulen machen keine Angaben im Bezug auf die erwartete Länge der Berichte.

Ebenfalls eingehend erklärt werden Zitier- und Bibliographierrichtlinien (21 Nennungen). Die Schulen liefern dazu meist beispielhafte Instruktionen, in denen das „Wie“ und das „Warum“ (Stichwort: Plagiat) des Zitierens und Bibliographierens beschrieben wird. Ausserdem machen mehr als die Hälfte der Schulen Angaben zum Aufbau des Gesamtberichts (19 Nennungen). Diesbezüglich herrschen jedoch grosse Unterschiede, die Angaben reichen von groben Strukturierungsvorgaben (in Titelblatt, Einleitung, Hauptteil, Schluss) bis hin zu beinahe erschöpfenden Kapiteleinteilungen.

Im Hinblick auf die formalen Vorgaben zeigen sich sprachregionale Unterschiede: Einerseits werden Vorgaben zu Formatierungen – die Bandbreite der Vorgaben reicht von Schriftgrösse und Titellayout bis zu ausführlichen Beispielen zur Textgestaltung (Textausrichtungen, Hervorhebungen, Fussnotenformatierungen etc.) – ausschliesslich in den Leitlinien der Deutschschweizer Gymnasien gemacht. Das Abstract oder die Zusammenfassung am Anfang der Maturaarbeit wird andererseits vor allem in der Westschweiz und im Tessin als obligatorischer Bestandteil der Arbeit gesehen.

Abb. VI.6: Formale Vorgaben (Anzahl Nennungen)



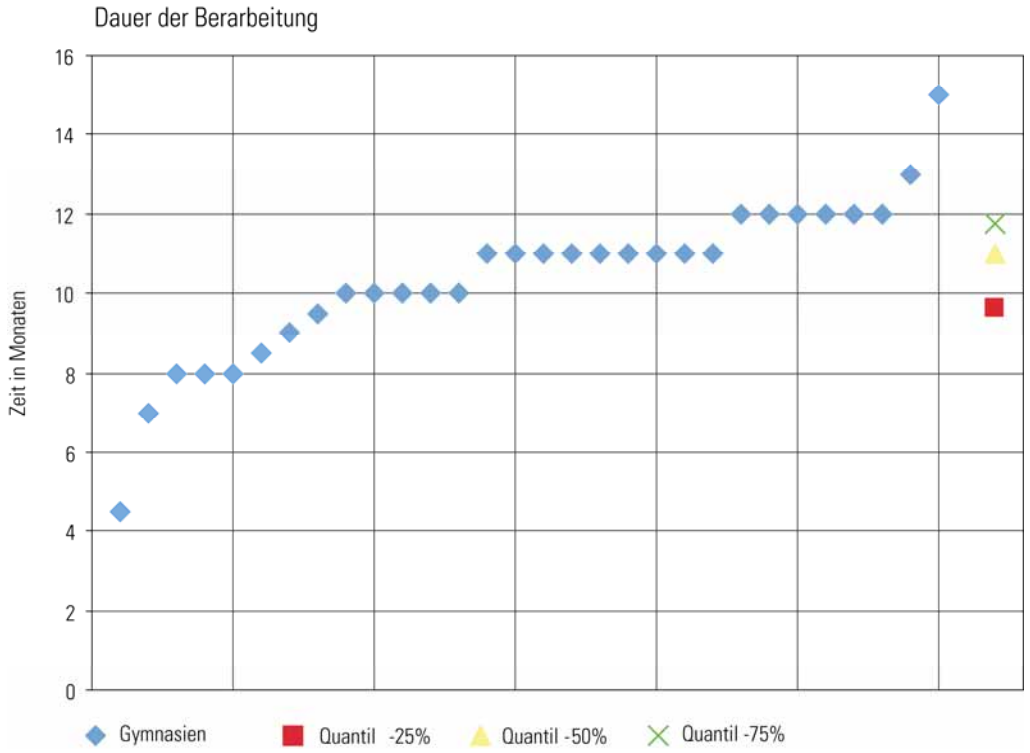
[50] Für Gruppenarbeiten gelten in der Regel entsprechend höhere Richtwerte.

3.2.5 Arbeitsprozess und Zeitpläne

Die Maturaarbeit wird mit Ausnahme einer Schule überall entweder im zweitletzten oder letzten Jahr vor Abschluss der Maturitätsschule erstellt. Der Arbeitsprozess gestaltet sich in der Regel so, dass nach einer Einführung (separate Veranstaltungen oder innerhalb des Fachunterrichts, vgl. Kap. 3.2.3 oben) die Themensuche und meist zeitgleich die Betreuungsfrage angegangen wird. Diese erste Phase erstreckt sich meist über mehrere Monate, kann aber auch innerhalb weniger Wochen stattfinden. Zur zweiten Phase zählen das Sammeln und Sichten von Daten, das Erstellen eines Feinkonzepts und in der Regel das Verfassen einer ersten Berichtsversion. Für diese Arbeiten sehen die Schulen vier bis acht Monate vor. Die Redaktion der Maturaarbeit erfolgt mehrheitlich in den letzten zwei Monaten vor dem Abgabetermin. Die mündliche Präsentation findet bis auf zwei Ausnahmen in allen Schulen erst nach der Annahme der Arbeit statt (teilweise unmittelbar oder bis zu zwei Monate später).

Der Zeitraum, welcher den Schülerinnen und Schülern zwischen definitiver Themenwahl und Abgabetermin für die Bearbeitung zur Verfügung steht, variiert mit einer Dauer von 4,5 bis 14,5 Monaten beträchtlich, das Gros der Schulen liegt zwischen neun (25% aller untersuchten Gymnasien) und elf (75%) Monaten jedoch nahe beieinander.

Abb. VI.7: Varianz der Bearbeitungszeit in Monaten



Den Schülerinnen und Schülern steht in etwa der Hälfte der Gymnasien der Deutschschweiz zusätzliche Unterrichtszeit für die Erarbeitung zur Verfügung (anstelle von Fachunterricht, zusätzlich zur Reduktion der Stundenzahl im allgemeinen Lehrplan der Maturitätsklassen). Anders präsentiert sich die Situation in den Gymnasien der lateinischen Schweiz: Aufgrund der Organisation im Seminarmodus fällt automatisch Unterrichtszeit für die Erarbeitung an.

Ein grosser Teil der Maturaarbeit wird aber ausserhalb des Unterrichts geleistet. Die Mehrheit der Schulen äussert sich dabei nicht zum erwarteten Zeitaufwand: Nur vier der 30 untersuchten Schulen geben den für die Redaktion und die Präsentation erwarteten Zeitaufwand approximativ an (zwischen 80 und 100 Arbeitsstunden). Insgesamt fällt aber auf, dass die Arbeitszeit der Maturandinnen und Maturanden sich auf eine lange Frist verteilt. Ein konzentriertes Arbeiten im Sinne einer längeren Blockarbeitszeit gibt es nur in seltenen Fällen.

3.2.6 Betreuung

Ein Grossteil der Deutschschweizer Gymnasien legt fest, dass zwischen den Schülerinnen und Schülern sowie den Betreuungspersonen ein Arbeitsvertrag abgeschlossen wird. In diesen Verträgen werden vorwiegend Thema und Grobkonzept sowie teilweise die Motivation und die Art der Zusammenarbeit zwischen Lehrenden und Lernenden festgehalten. Eine solche Projektvereinbarung gilt als Grundlage für Folgearbeiten und dient in einigen Fällen auch als Vergleichsbasis für die spätere Beurteilung der Leistungen der Schülerinnen und Schüler. Sie ist aber nicht gleichzusetzen mit dem Arbeitsjournal, das als Begleit- und Prozessinstrument in allen Schulen eingeführt wurde und das die einzelnen Schritte des Arbeitsprozesses dokumentiert und reflektiert. In einigen wenigen Gymnasien wird der dokumentierte Arbeitsprozess als obligates Teilkapitel in die Maturaarbeit integriert. Die Reflexion des Arbeitsprozesses wird in den meisten Schulen bei der Abschlusszertifizierung der Maturaarbeit mitberücksichtigt.

Die Betreuung geschieht in der Regel durch eine oder mehrere Lehrpersonen, wobei das Suchprimat zumeist bei den Schülerinnen und Schülern liegt. In einem kleineren Teil der Schulen ist zudem eine (Co-)Betreuung durch externe Experten möglich. Zu den selteneren Gestaltungsvarianten gehört auch, dass Schulleitungsmitglieder als Expertinnen und Experten für die Abschlussevaluation beigezogen werden. Die Betreuungsanfragen – meist schriftlicher Art in Form von erweiterten Themenfeldbeschreibungen – werden von den Lehrpersonen begutachtet und obliegen deren expliziter Zustimmung. Bemerkenswert ist diesbezüglich, dass nur in wenigen Dokumenten die spezifisch fachliche Qualifikation der betreuenden Personen als conditio sine qua non für eine mögliche Betreuung gilt. Zu vermuten ist deshalb, dass neben der fachlichen Zuständigkeit und der zeitlichen Verfügbarkeit auch die persönliche Affinität zwischen Schülerinnen und Schülern auf der einen und den Lehrpersonen auf der anderen Seite bei der Auswahl von Bedeutung ist. Der Betreuungsrhythmus bzw. die minimale Häufigkeit des Austauschs zwischen den Lernenden und den betreuenden Lehrpersonen wird nur in wenigen Fällen (neun Schulen) explizit festgehalten.

Die schriftlich festgehaltenen Pflichten der betreuenden Lehrpersonen unterscheiden sich zwischen den untersuchten Schulen nur wenig. Sie sind vor allem auf die inhaltliche und methodische Unterstützung ausgerichtet. Die Rolle der betreuenden Personen wird jedoch durchaus unterschiedlich interpretiert und reicht von Beratung und Beurteilung über „geistige Geburtshilfe“ bis hin zu Überwachung und Disziplinierung. Die Mehrheit der Schulen formuliert dabei eine Mittelposition zwischen formativer und zertifikativer Begleitung.

Eine ausgewiesene Einführung in die Betreuungsarbeit (als Informationsveranstaltung oder als schriftliche Dokumentation) ist nur in einem kleinen Teil der Deutschschweizer Gymnasien vorgesehen, in den Gymnasien der Romandie und des Tessins sind keine spezifischen Vorbereitungen dokumentiert. Den betreuenden Lehrpersonen stehen – wie auch den Schülerinnen und Schülern – die Leitfäden zur Maturaarbeit zur Verfügung, in denen mehr oder weniger konkret festgehalten wird, welche Erwartungen und Zielsetzungen mit der Maturaarbeit verbunden sind.

Es ist davon auszugehen, dass die Schulen in der Regel die allgemeine Ausbildung der Lehrpersonen als genügend erachten, um eine adäquate Betreuungsarbeit leisten zu können. Eine auf den disziplinären Hintergrund der jeweiligen Maturaarbeit ausgerichtete professionelle Kompetenz wird jedenfalls nur in einigen wenigen Schulen explizit vorausgesetzt. Im Zweifelsfall kann ein solcher Verzicht auf eine fachlich adäquate Betreuung aber die Gefahr einer minderen Betreuungsqualität auf der sachlichen Ebene beinhalten.

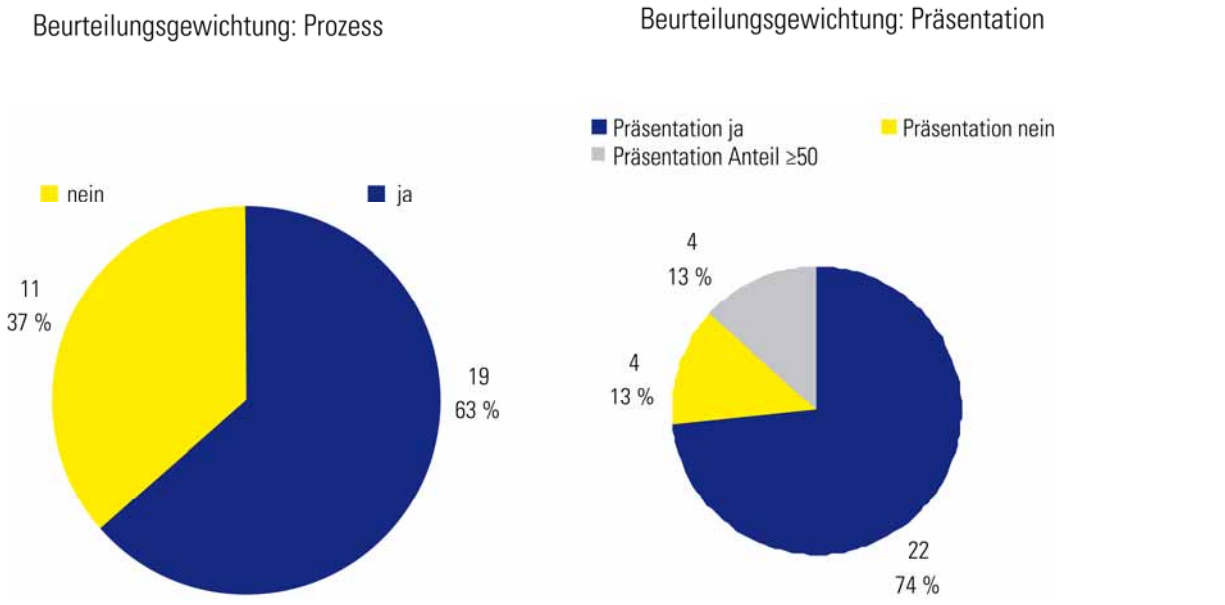
Die Betreuungsleistung wird in den überwiegenden Fällen als Pflicht der Lehrpersonen gefasst, wobei die maximale Anzahl von zu betreuenden Arbeiten zwischen drei und acht schwankt. Die Entlastung für Betreuungsarbeiten ist nur bei acht Schulen explizit geregelt. Sie geschieht in der Regel in Form von Unterrichtsentlastung oder eines Eintrags im Pensenblatt.

3.2.7 Beurteilung und Bewertung

Im Bezug auf die Beurteilungskriterien sind drei Varianten auszumachen. Ein Teil der Schulen (darunter die Hälfte der Deutschschweizer Gymnasien) stellt für die Maturaarbeiten sämtlicher Fachbereiche allgemeine Evaluationskriterien auf: In den Wegleitungen werden stichwortartig allgemeine Bewertungskriterien zur Qualität von Idee, Konzept und Umsetzung, sowie der Präsentation aufgeführt. Eine zweite Gruppe von Gymnasien konkretisiert die Kriterien anhand fachlicher Bestimmungen. So entstehen umfangreiche Beurteilungsschemata, die in konkrete Beurteilungsfragen münden. Fasst man die Beurteilungs- und Bewertungskriterien dieser beider Gruppen inhaltlich-kategorial zusammen, lassen sich auf der Ebene der Einzelschule materiell kaum Unterschiede feststellen. So werden die Arbeiten hinsichtlich Form und Ausdruck, Inhalt und Argumentationsgang sowie Methodeneinsatz beurteilt. Eine dritte Gruppe von Gymnasien (der überwiegende Teil der Westschweizer und Tessiner Gymnasien) überlässt die Definition der Beurteilungskriterien den betreuenden Lehrpersonen. Ob und in welchem Ausmass diese fach- und lehrpersonenbezogenen Beurteilungskriterien den Schülerinnen und Schülern kommuniziert werden, ist indes nicht weiter beschrieben.

Weitere Unterschiede zwischen den Schulen ergeben sich in Bezug auf die der Benotung zugrunde liegende Gewichtung der einzelnen Arbeitsschritte und Produkte (vgl. Abbildung VI.8). Unterschiedliche Gewichtungen erfährt das Verhältnis von schriftlicher Arbeit oder Dokumentation auf der einen und der mündlichen Präsentation auf der anderen Seite. Der Anteil der Präsentation an der Gesamtnote kann dabei bis zu zwei Dritteln ausmachen.

Abb. VI.8: Gewichtung der Abschlussbeurteilung



Die meisten Schulen weisen ebenfalls darauf hin, dass auch der eigentliche Arbeitsprozess beurteilt wird. Die Beurteilung des Arbeitsprozesses – darunter fallen die Arbeitsmotivation und Arbeitshaltung, die Einhaltung von Terminen sowie Elemente des selbstverantwortlichen Lernens – fliesst bei der Mehrheit der Schulen (63%) in die Bewertung ein. Teilweise wird hier auch die Selbsteinschätzung der Schüler und Schülerinnen einbezogen. Zudem zählt in vier Gymnasien das Konzept als expliziter Bestandteil der Bewertung.

Für Gruppenarbeiten verlangen die meisten Schulen, dass die individuellen Anteile erkennbar und somit jeweils einer Autorin oder einem Autor zuzurechnen sind. Abweichend davon kann mitunter das schriftliche Produkt nach Vereinbarung als Teamarbeit bewertet werden, während die Präsentationen in der Regel für jedes Teammitglied individuell bewertet werden.

Ungenügende Noten haben unterschiedliche Wirkungen: In einem überwiegenden Teil der Gymnasien gilt eine genügende Gesamtnote für die Maturaarbeit und deren Präsentation als Zugangskriterium für die Zulassung zu den Maturaprüfungen. Als Vorstufe davon gilt in einem kleineren Teil der Gymnasien eine genügende schriftliche Maturaarbeit auch als Voraussetzung für die Zulassung zur Präsentation, in anderen Schulen findet eine Präsentation auf jeden Fall statt.

Das MAR 95 schreibt Titel- und Noteneintrag vor, ohne dass die Note als promotionswirksam gelten muss. In der Umsetzung sind verschiedene Vorgehen zu finden: In der Regel finden sowohl Thema wie Prädikat Erwähnung im Maturazeugnis. In einer einzigen Kantonsschule fliesst die Note für die dokumentierte Arbeit in eine fachspezifische Note (gemäss Fachzuordnung der Arbeit) ein. Keine eigentliche Benotung, sondern Prädikate kennen vor allem die Gymnasien der Westschweiz.

3.3 Zwischenbilanz

Für die Beurteilung der Rahmenbedingungen, unter welchen die Maturaarbeiten verfasst werden, liegt eine Vielzahl von teilweise umfangreichen Papieren vor. Gleichwohl befinden sich viele Schulen in einem Prozess, der darauf abzielt, die Richtlinien weiter auszuarbeiten, zu verfeinern und zu erproben. Die laufende Veränderung der Grundlagen macht eine Bestandesaufnahme zwar schwierig, jedoch nicht unmöglich. Für die hier vorgelegte Analyse der dokumentierten[51] Rahmenbedingungen gilt, dass in Entsprechung zur Projektphase der evaluierten Maturaarbeiten die Vorgaben für das Schuljahr 2006/07 ausschlaggebend sind.

Zusammenfassend kann gesagt werden, dass die Dokumentation der Verantwortlichkeiten, Rahmenbedingungen und Arbeitsprozesse von Schule zu Schule sehr unterschiedlich sind. Die Verschiedenheit in den Zielvorstellungen und den Realisierungen einer vermeintlich einfachen Aufgabenstellung (vgl. nationale Vorgaben in Kap. 3.1.1) betrifft sowohl die formale Ebene (welche Elemente werden geregelt?), als auch die inhaltliche (welche Ziele und Vorgaben werden formuliert?). Zentral erscheinen diesbezüglich folgende Merkmale:

- **Richtlinien** für die Schülerinnen und Schüler sind in allen untersuchten Schulen vorhanden. Die inhaltliche Breite dieser Anleitungen lässt jedoch vermuten, dass diese nicht in allen Schulen denselben Stellenwert besitzen: Schmalere Richtlinien lassen sowohl den Betreuenden als auch den Schülerinnen und Schülern mehr Spielraum für die spezifische Ausgestaltung der Projekte. Ausführliche Beschreibungen und detaillierte Beispieltexte sind eher dafür nutzbar, den Schülerinnen und Schülern konkrete Hilfestellungen zu geben und gleichzeitig die betreuenden Lehrpersonen von einer formalen Kontrollarbeit zu entlasten. Die objektivierten Anforderungen gewährleisten nicht zuletzt Transparenz für das Verfahren und geben Normen für rollengerechtes Verhalten der Lehrenden und der Lernenden vor.
- **Organisatorisch** sind sprachregionale Unterschiede festzustellen: Die Gymnasien der Kantone Fribourg, Tessin, Waadt und Wallis organisieren den Unterricht um die Maturaarbeit als so genannte Seminarveranstaltung. Innerhalb dieser Unterrichtsgefässe werden Themen behandelt, die üblicherweise die Grenzen der gymnasialen Schulfächer sprengen und somit als Gesamtrahmen der Interdisziplinarität Vorschub leisten. Die verfassten Maturaarbeiten weisen somit zwar eine Themenbindung auf, jedoch kaum eine Themenbeschränkung. Die typische Organisationsform in den Deutschschweizer Gymnasien hingegen fusst auf dem Prinzip der maximalen individuellen Motivation der Schülerinnen und Schüler. Ihnen steht in der Regel sowohl die Themen- als auch die Betreuendenwahl offen. Weder von der einen noch der anderen Organisationsform abhängig ist die Frage der fachspezifischen Betreuung durch die Lehrpersonen und die tatsächliche Einschränkung der Themenwahl. Die organisatorischen Unterschiede sind demnach weniger eine Antwort auf unterschiedliche pädagogische Fragen, sondern vielmehr Ausdruck der Ressourcenbindung einer Schule.

[51] Hier sei nochmals betont, dass nicht alle Kooperationsabsprachen durch die Schule dokumentiert werden. Solche informellen Rahmenabsprachen entziehen sich der Analyse im Rahmen dieser Untersuchung.

- Das Einüben **wissenschaftlichen Arbeitens** steht als Ziel in beinahe allen Dokumentationen. Dass mit der Maturaarbeit also ein wissenschaftspropädeutischer Beitrag geleistet werden soll, der die Arbeitsweise von weiterführenden Bildungseinrichtungen im Tertiärbereich simuliert, ist mithin unbestritten. Umso mehr erstaunt die Tatsache, dass die wenigsten Schulen eine fach- oder fachbereichsadäquate Ausbildung der Betreuungspersonen voraussetzen: Gemäss den schuleigenen Vorgaben ist zumindest nicht ausgeschlossen, dass alle Lehrpersonen alle Arten von Arbeiten betreuen können. Wie stark die Professionalität der Betreuung durch nicht dokumentierte Konventionen gesichert ist, kann aus den verfügbaren Dokumenten nicht eindeutig erschlossen werden.
- Bezüglich der **Leistungsmessung** fällt auf, dass im Vergleich zu unteren Schulstufen der Volksschule, die Beurteilungspraxis für die Maturaarbeit wenig Aufmerksamkeit erfährt. Die Benotungsbasis präsentiert sich dementsprechend höchst unterschiedlich. So fliessen teils formative Elemente in die summative Evaluation mit ein, an einigen Schulen wiederum beträgt der Anteil der Präsentation über 50% der Abschlussnote. Es stellt sich daher die Frage, was genau unter dem Gesichtspunkt der Wissenschaftspropädeutik benotet werden soll: Ist es das Engagement und die Motivation, der erzielte Fortschritt im Arbeitsprozess, das schriftlich vorgelegte Resultat oder dessen mündliche Präsentation? Ob die von Bund und Kantonen beschlossen Änderungen bezüglich des Stellenwerts der Maturaarbeit als Abschlussarbeit die notwendigen Klärungen bringen werden, bleibt offen.
- Für die Unterschiede und Gemeinsamkeiten auf der **formalen Ebene** ist festzuhalten, dass den Kantonen einheitliche nationale Vorgaben fehlen. Dieses Fehlen einer Regelung wird jedoch oft von den Kantonen nicht aufgefangen und kompensiert: Kaum ein im Untersuchungssample vertretener Kanton macht einheitliche und schulübergreifende Vorgaben zu Zielen, Inhalten und Abläufen bezüglich der Maturaarbeiten. Dieser Umstand belässt den einzelnen Schulen viel Spielraum, die Abläufe selber zu regeln oder aber auf eine dokumentierte Regelung zu verzichten. Das Fehlen von dokumentierten formalen Regelungen in einzelnen Schulen soll nicht suggerieren, dass die jeweiligen Prozesse unbestimmt bleiben – es ist zu vermuten, dass etliche Vorgaben direkt „am Ort des Geschehens“, nämlich im alltäglichen Unterricht erarbeitet werden. Die grossen Unterschiede bezüglich der formalen Ebene offenbaren jedoch, dass über der Regelungsbedarf für das Erarbeiten von Maturaarbeiten keine Einigkeit besteht, und zwar weder auf nationaler noch auf kantonaler Ebene. Bei der Ausgestaltung der Rahmenbedingungen führt dies zu einem nicht unerheblichen Ermessensspielraum für die Gymnasien und für die betreuenden Lehrpersonen, dessen Nutzung zu grossen Differenzen bei den Maturaarbeiten als Produkten eines veranstalteten Lernprozesses führen kann. Als Desiderat ist hier eine längsschnittliche Analyse der Maturaarbeiten einzufordern, die zu zeigen hätte, ob die Regelungsdichte einzelner Schulen Auswirkungen auf die Gleichwertigkeit der Arbeiten innerhalb einer Abschlussklasse oder eines Fachbereichs einerseits, sowie andererseits auch auf die Arbeiten über die Jahrgänge und sogar Kantonsgrenzen hinweg hat. Natürlich sind schriftlich dokumentierte Vorgaben nur dann von Nutzen, wenn die Richtlinien nicht bloss als „Papiertiger“ enden, sondern auch tatsächlich eingesetzt werden, um die Rahmenbedingungen der Maturaarbeiten innerhalb einer Schule zu vereinheitlichen.

4 Die Maturaarbeit im Urteil der Schülerinnen und Schüler

Die Befragung von Maturandinnen und Maturanden stellt ein wesentliches Element der Analyse der Maturaarbeiten dar. Durch ihre Antworten können einerseits elementare soziodemografische Informationen zum Untersuchungssample gewonnen werden, andererseits erlaubt es die Befragung auch, subjektive Einschätzungen der Schülerinnen und Schüler zu verschiedenen Aspekten der gymnasialen Abschlussprojekte zu erhalten.

Bereits im Projekt EVAMAR I fand eine Befragung von Schülerinnen und Schülern statt. Obwohl der Fokus von EVAMAR I eher auf systemischen Fragen lag, wurden im Rahmen des Projekts 2 zu den „Objectifs pédagogiques transversaux“ auch Fragen zur Maturaarbeit, insbesondere zu den Bereichen *Themenwahl*, *Betreuung* und *Realisierungsbedingungen* gestellt. Aus diesem Grund wurde bei der Auswahl der Fragen nach Möglichkeit ein Bezug zu

EVAMAR I hergestellt, um einen Quasi-Längsschnitt zeichnen zu können. So wurden Items zum Fachbezug, zur Themenwahl, zur Betreuung und zu den Realisierungsbedingungen unverändert oder leicht modifiziert übernommen.

In Übereinstimmung mit dem in Kapitel 2.1 dargestellten Wirkungsmodell wurden im Schülerinnen- und Schülerfragebogen Aspekte individueller Einflüsse auf die Qualität der Maturaarbeit erfragt. Diese Aspekte erstrecken sich retrospektiv auf die Vorgeschichte, den Projektverlauf und die schulische Bewertung des Resultats, aber auch prospektiv auf den Wert des Lernerfolgs für weitere Qualifikationsprozesse und im Hinblick auf die persönliche Entwicklung. Im Einzelnen handelt es sich um folgende Fragebereiche:

- soziodemographische Angaben
- Fachbezug der Maturaarbeit
- Themenwahl
- Beurteilung der Maturaarbeit
- Betreuung
- Unterstützung und Ressourcen
- Einschätzung von Nutzen und Aufwand der Maturaarbeit.

Die folgenden Abschnitte nehmen Bezug auf diese Aspekte und stellen die Ergebnisse der Befragung in deskriptiver Form und in ersten analytischen Ausführungen dar, die im folgenden Kapitel anhand des Ratings weiter vertieft werden.

4.1 Fächerwahl

Die von den Schülerinnen und Schülern belegten Schwerpunktfächer liegen zu 31% im erst- und fremdsprachlichen Bereich, zu 21,9% im Bereich Naturwissenschaften, zu 31,7% im geistes- und sozialwissenschaftlichen Bereich sowie zu 15,4% im Bereich Kunst, Musik und Sport. Im Vergleich zu den Ergebnissen aus der Studie EVAMAR I zeigen sich hinsichtlich der Fächerwahl leichte Unterschiede. Der Bereich Erst-/Fremdsprache wurde von den im Rahmen von EVAMAR I befragten Maturandinnen und Maturanden zu 36% als Schwerpunktfach gewählt, geistes- und sozialwissenschaftliche Arbeiten machten 27% aus, auf die Naturwissenschaften fielen 25,8% und 11,1% der Befragten beschäftigten sich in ihrem Schwerpunktfach mit einem Thema aus den Gebieten Kunst, Musik oder Sport (Ramseier et al. 2005, S. 67).[52] Nach einer Gewichtung ergeben sich für die hier vorliegenden Daten die folgenden Prozentverteilungen: Erst-/Fremdsprache 31,6%, naturwissenschaftliche Fächer 23,3%, geistes- und sozialwissenschaftliche Fächer 30,8% und musisch-künstlerische Fächer sowie Sport 14,3%. Es kann also festgestellt werden, dass sich die Schwerpunktwahl im vorliegenden Sample leicht zugunsten des Bereichs Gestaltung und Bewegung sowie geistes- und sozialwissenschaftlicher Fächer verschoben hat, während die erst- und fremdsprachlichen Fächer und die Naturwissenschaften seltener gewählt wurden. Eine etwas andere Verteilung ergibt sich hinsichtlich des Ergänzungsfachs. Hier liegen die geistes- und sozialwissenschaftlichen Fächer mit 54,7% deutlich vorne, auf den naturwissenschaftlichen Bereich entfallen 16,9%, auf die künstlerisch-musischen Fächer sowie Sport 28,4%.

4.2 Fachbezug der Maturaarbeit und Themenwahl

Im Zusammenhang mit der Fächerwahl der Maturandinnen und Maturanden interessierte ebenfalls die Zuordnung der eigenen Maturaarbeit zu einem oder mehrerer Fächer. Gefragt wurde sowohl nach einer primären

[52] Ein direkter Vergleich der Ergebnisse aus beiden Studien ist streng genommen nicht zulässig, da die Zusammensetzung der Stichproben und der Erhebungszeitraum im Schuljahr erheblich voneinander abweichen. Hinzu kommt, dass die Daten aus EVAMAR I durchgängig nach Sprachregion und Geschlecht gewichtet wurden, während die hier dargestellten Ergebnisse zumeist auf ungewichteten Daten beruhen. Werden aber die hier vorliegenden Daten ebenfalls nach Sprachregion und Geschlecht gewichtet, so bleiben die genannten Unterschiede zu EVAMAR I bestehen.

(Hauptfach-) als auch nach einer sekundären (Nebenfach-) Zuordnung, wobei in beiden Fällen explizit auch Mehrfachnennungen zugelassen wurden.

Die Mehrheit der Schülerinnen und Schüler (66,4%) ordnet sich dabei nur einem einzigen Hauptfach zu. Aber es gibt auch eine Reihe von Personen, die bis zu vier oder mehr Hauptfächer nennen. Insgesamt geben 20,8% der Maturandinnen und Maturanden mindestens zwei Hauptfächer an, es gibt aber auch 12,8% die keine Zuordnung zu einem Hauptfach vorgenommen haben. 7,9% ordnen ihre Maturaarbeit weder einem Haupt- noch einem Nebenfach zu und 49,9% geben nur entweder ein Haupt- oder ein Nebenfach an, in dem sie ihre Arbeit verorten würden. Hierin sind auch diejenigen Aussagen enthalten, bei denen dasselbe Fach als Haupt- und als Nebenfach genannt wurde. Daraus kann man schliessen, dass nach den Angaben der Maturandinnen und Maturanden knapp 50% der Maturaarbeiten monodisziplinär ausgerichtet sind. 28% der Befragten geben an, ihre Maturaarbeit in zwei Fächern zu verorten, wobei dies Haupt- und Nebenfächer sein können. Weitere 14,1% nennen mehr als zwei Haupt- oder Nebenfächer, denen sie die Arbeit zuordnen.

Diese Ergebnisse unterscheiden sich erheblich von denen aus der ersten EVAMAR-Studie. Dort wurden die Schülerinnen und Schüler gefragt, worauf sich ihre Maturaarbeit bezieht. Die Antwortalternativen waren mit *Auf ein einziges Schulfach*, *Auf mehrere Schulfächer* und *Auf ein Thema ausserhalb der regulären Schulfächer* vorgegeben. Je 38% der Schülerinnen und Schüler gaben an, sich auf ein einziges Fach oder auf ein Thema ausserhalb der regulären Schulfächer zu beziehen, nur 24% verorteten ihre Arbeit zwischen mehreren Schulfächern. Die Gewichtung scheint in dieser Frage keine Rolle zu spielen, da die Ergebnisse aus der hier vorliegenden Studie sich auch mit einer Gewichtung nach Sprachregionen und Geschlecht kaum verändern. Wichtiger scheint hingegen die Art der Fragestellung zu sein, die tatsächlich einen Vergleich zwischen den beiden Studien behindert. Vermutlich haben sich viele der Befragten durch die vorgegebene Auswahlliste der Themen eher dazu aufgefordert gefühlt, eine Zuordnung zu finden, als dies bei EVAMAR I der Fall war. Daraus könnte sich erklären, dass nur 7,9% der Befragten ihre Arbeit nicht zu einem der Fachgebiete zuordnen konnten.

Neben der Frage, ob die Maturaarbeit als mono- oder interdisziplinär eingestuft werden kann, interessiert uns ebenfalls in welchem Bereich die Maturandinnen und Maturanden ihre Maturaarbeiten ansiedeln. Dafür sind die Angaben zur fachlichen Spezifikation der Arbeiten genauer zu analysieren: Tabelle VI.11 listet auf, wie die Befragten ihre Maturaarbeit in das Feld übergreifender Fachgebiete einordnen.[53] Die Fächerzuordnungen Deutsch, Französisch, Italienisch, Englisch, Spanisch, Russisch, Latein und Griechisch werden hier zum Bereich Erst-/Fremdsprache zusammengefasst. Mathematik, Anwendungen der Mathematik, Physik, Chemie und Biologie sowie Informatik bilden den naturwissenschaftlichen Bereich. Der geistes- und sozialwissenschaftliche Bereich erstreckt sich auf Geschichte, Geographie, Philosophie, Wirtschaft, Staatskunde, Recht, Pädagogik, Psychologie, Soziologie und Religion. Der Bereich Gestaltung und Bewegung umfasst Bildnerisches Gestalten, Kunstgeschichte, Theater, Musik und Sport.[54]

Tab. VI.11: Themenwahl nach übergreifendem Fachgebiet (Angaben in Prozent, Mehrfachantworten möglich)

Fachgebiet	Prozent der Maturaarbeiten
Erst- und fremdsprachlicher Bereich	30,9
Naturwissenschaftlicher Bereich	21,8
Geistes- und Sozialwissenschaftlicher Bereich	31,5
Gestaltung und Bewegung	15,3

[53] Dabei können im Fall von interdisziplinären Arbeiten einzelne Schülerinnen und Schüler auch mehrfach in der Darstellung auftauchen, andere, die keine Zuordnung vorgenommen haben tauchen nicht auf.

[54] Diese Selbstverortungen sind nicht unbedingt mit den curricularen Fächerbezeichnungen kongruent. So kommt etwa der Soziologie nicht der Stellenwert eines Schulfachs zu.

Von den 802 Zuordnungen, die von den Autorinnen und Autoren der Maturaarbeiten zu je mindestens einer Domäne gemacht wurden, sind 86,9% in genau einem der vier Fachgebiete anzusiedeln. Innerhalb der Fachgebiete gibt es dabei teilweise von einzelnen Personen Angaben zu mehreren Disziplinen, wobei dies im Bereich der Geistes- und Sozialwissenschaften am häufigsten vorkommt: Gut ein Fünftel dieser Maturaarbeiten können nach Sicht der Verfasserinnen und Verfasser auf mehr als eine Disziplin innerhalb dieses Bereichs bezogen werden. Bei den naturwissenschaftlichen Arbeiten sind dies nur 15,5%, im künstlerisch-musischen Bereich 5,5% und im sprachlichen Bereich lediglich 3,2%. Sicherlich ist dies auch durch die unterschiedliche Anzahl von Disziplinen und damit der unterschiedlichen Kombinationsmöglichkeiten in den jeweiligen Bereichen zu erklären, innerhalb des sprachlichen Bereichs ist jedoch eher eine monodisziplinäre Ausrichtung auffällig. Es gibt zudem eine Reihe von Zuordnungen, die über die Fachgebietskategorien hinausgehen. Bei 13,1% der Zuordnungen handelt es sich um fachgebietsübergreifende Mehrfachnennungen, die einen starken Anspruch auf Interdisziplinarität zum Ausdruck bringen.

Die Maturandinnen und Maturanden haben im Laufe ihrer Schulzeit überwiegend fachgebundene und monodisziplinäre Lernprozesse durchlaufen. Ein hoher Anteil an Arbeiten, die in Verbindung mit mehreren Fächern stehen, wäre aus diesem Grunde nicht zu erwarten. Dennoch ist auffällig, wie selten Maturaarbeiten als interdisziplinär deklariert werden. Eine mögliche Erklärung könnte in den Rahmenbedingungen liegen, da diese zeigen, dass Interdisziplinarität nur in Schulen aus der französischsprachigen Schweiz explizit verlangt wird (vgl. Kapitel 3).

In der Romandie gibt es hinsichtlich des Anteils der monodisziplinären Arbeiten keinen Unterschied zu den anderen Sprachregionen, der Anteil interdisziplinärer Arbeiten liegt hingegen tatsächlich mit 35,1% etwas höher, wobei dafür der Anteil der Arbeiten ohne Fachzuordnung niedriger ist. Dies kann daran liegen, dass sich die französischsprachigen Maturandinnen und Maturanden aufgrund der vorgegebenen Rahmenbedingungen eher mit der Option der Interdisziplinarität explizit auseinandergesetzt haben und deshalb eher eine Zuordnung zu mehreren Fächern vornehmen.

4.3 Themenwahl bei der Maturaarbeit

Rund drei Viertel der Befragten geben an, das Thema ihrer Maturaarbeit frei bestimmt zu haben, knapp 14% haben ihre Wahl innerhalb eines vorgegebenen Rahmenthemas vorgenommen und 9% haben aus einer Liste mit Vorschlägen wählen können. Vergleicht man diese Angaben mit den von den Schulen vorgegebenen Rahmenbedingungen, zeigt sich, dass die Maturandinnen und Maturanden möglicherweise auch eine themen- oder schulfachgebundene Wahl als freie Wahl ansehen. Nur 35% der Schulen ermöglichen nämlich durch ihre Rahmenvorgaben eine freie Auswahl des Themas, aber weitere 47% erlauben eine durch ein Thema oder ein Schulfach eingeschränkte Wahl. Lediglich 6% der Schulen geben eine Liste zur Auswahl vor.

Der Regionalvergleich zeigt, dass die Einschätzung der Wahlmöglichkeiten durch die Schülerinnen und Schüler in den drei Sprachregionen recht unterschiedlich ausfallen, wobei die freie Themenwahl in der Deutschschweiz dominiert. In der Romandie wird die Wahl häufiger über ein Rahmenthema oder eine Liste mit Vorschlägen bestimmt. Im Kanton Tessin halten sich die freie Themenwahl und die Vorgabe eines Rahmenthemas annähernd die Waage (siehe Tabelle VI.12).

Tab. VI.12: Wahl des Themas für die Maturarbeit nach Sprachregion (Angaben in Prozent)

Art der Themenwahl	Deutschschweiz	Romandie	Tessin	Gesamt
Freie Themenwahl	94,8	45,9	43,6	75,9
Aus Rahmenthema	3	29,1	40	13,7
Liste mit Vorschlägen	1,1	23,3	14,5	9,1
Andere	1,1	1,7	1,8	1,3

Tatsächlich kennen gemäss den Rahmenbedingungen alle Gymnasien der Deutschschweiz die freie Themenwahl. Nur bei einem Viertel dieser Schulen besteht als zusätzlicher Wahlmodus die Möglichkeit, die Themen anhand von Listen auszuwählen. Auffällig ist aber die Differenz der Einschätzungen der Schülerinnen und Schüler zu den schulischen Rahmenbedingungen in den Tessiner Gymnasien, bei denen die Maturaarbeit in Seminarform organisiert wird: Die Schülerinnen und Schüler können ein zu einem bestimmten Rahmenthema angebotenes Seminar belegen und aus diesem Rahmenthema wiederum ein Thema für die Maturaarbeit wählen. Insofern wird die Wahl durch das Veranstaltungsangebot limitiert. Von den Maturandinnen und Maturanden wird dieser Umstand jedoch offensichtlich nicht als Einschränkung angesehen, da immerhin noch 43,6% der Befragten angeben, ihr Thema frei gewählt zu haben. In der französischsprachigen Schweiz gibt es je nach Kanton unterschiedliche Rahmenvorgaben, nach denen teils frei gewählt werden kann, teils aus einer Liste von Themen und teils aus einem Rahmenthema.

Vor diesem Hintergrund lassen sich vermutlich auch die Differenzen zur ersten EVAMAR-Studie erklären nach der nur 8% der Tessiner Schülerinnen und Schüler ihr Thema frei wählen können. Weiterhin wurde im Rahmen von EVAMAR I festgestellt, dass 65% der französischsprachigen Maturandinnen und Maturanden angehalten sind, ihr Thema aus einer Liste bzw. aus einem Rahmenthema zu wählen. Hier scheint also eine Änderung der Wahlmodi stattgefunden zu haben.

Bei der Festlegung eines zu bearbeitenden Themas ist in neun von zehn Fällen das eigene Interesse von Gewicht. Dies ist nicht weiter erstaunlich, kann man sich doch nur schwer vorstellen, dass Maturandinnen und Maturanden über längere Zeit selbstständig ein Thema bearbeiten, das sie nicht oder nur marginal interessiert. Rund 30% der Befragten geben an, dass sie sich an der Einschätzung der eigenen Vorkenntnisse orientiert haben; ein Viertel hebt die Herausforderung durch das Thema hervor. Knapp 20% verweisen bei den Einflüssen auf ihre Studienabsicht. Lediglich jeder oder jede Zehnte gibt an, dass die Option, die Arbeit in einer bestimmten Teamkonstellation verfassen zu können, ein Beweggrund für die Themenwahl war. Weiter haben Ratschläge von Bekannten und Eltern ebenso einen untergeordneten Stellenwert wie der Bezug zu einer Lehrperson. Kaum eine Entscheidungsgrundlage stellt die antizipierte Leichtigkeit der Ausführung des Projekts dar. Nur 3,2% der Befragten gaben an, dass sie von dem Motiv beeinflusst wurden, eine Arbeit zu schreiben, die bei der Ausführung möglichst geringe Anforderungen beinhaltet.

Tab. VI.13: Motive für die Themenwahl (Angaben in Prozent, Mehr-fachantworten möglich)

Motiv	Prozent der Befragten
Interesse	92,4
Vorkenntnisse	29,8
Herausforderung	24,5
Studienabsicht	18,9
Teamorientierung	9,9
Ratschlag LP	6,7
Ratschlag Eltern	6,4
Ratschlag von Bekannten	6,2
Betreuungsperson	6,1
Leichte Ausführbarkeit	3,2

Die Ergebnisse aus EVAMAR I fallen ähnlich wie die hier beschriebenen aus. 90% der damals befragten Maturandinnen und Maturanden gaben das eigene Interesse als Motiv für die Themenwahl an und 34% betonten ihre

Vorkenntnisse auf dem Gebiet der Maturaarbeit. Jeweils weniger als 10% der Befragten konnten bestätigen, dass sie durch die vermutete Einfachheit der Realisierung oder die Ratschläge von Eltern, Lehrpersonen oder Kolleginnen und Kollegen beeinflusst wurden (Pagnossin, Aliata & Dozio, 2005 S. 185).

Dass das Interesse der Schülerinnen und Schüler bei der Wahl des Themas ausschlaggebend ist, kann nicht sonderlich überraschen. Selbst bei limitierenden Rahmenbedingungen haben die Maturandinnen und Maturanden oft den Eindruck, dass die Themenwahl durch ihre eigenen Präferenzen bestimmt wird. Die Orientierung an Vorkenntnissen und die Suche nach Herausforderungen können so auch als Elemente gelesen werden, die das Interesse zusätzlich steigert. Die Orientierung an einer Studienabsicht wiederum scheint auf einer funktionalistischen Überlegung zu gründen, wobei dieses Kalkül der individuellen Nutzenmaximierung nicht dazu führt, die Maturarbeit danach auszuwählen, wie leicht sie auszuführen scheint. Aus den Angaben wird deutlich, dass die Maturandinnen und Maturanden vorwiegend ein intrinsisches Interesse bezeugen, sich ernsthaft mit der Thematik ihres Projekts auseinanderzusetzen.

4.4 Bewertung der Maturaarbeit

Für ihre Maturaarbeit haben rund 86% der Schülerinnen und Schüler eine gute oder sehr gute Note erhalten; bei 12,5% lautet die Bewertung "genügend". Lediglich 1,3% der Befragten erhielten eine ungenügende Bewertung. In diesem Zusammenhang ist darauf hinzuweisen, dass die Bewertungskriterien von Kanton zu Kanton und teilweise auch von Schule zu Schule variieren: Die Note bezieht sich häufig nicht allein auf das schriftliche Produkt, sondern auch auf die mündliche Präsentation der Ergebnisse. Hinzu kommt, dass die Relation von schriftlichen und mündlichen Bewertungskomponenten unterschiedlich ausgeprägt ist. Ebenso fallen die Konsequenzen bei Maturaarbeiten mit ungenügender Note schulspezifisch unterschiedlich aus (vgl. dazu Kapitel 3).

Tab. VI.14: Benotung der Maturaarbeit (Angaben in Prozent)

Note	Prozent der Maturaarbeiten
sehr gut, très bien, Noten 6 und 5,5	46,5
gut, bien, Noten 5 und 4,5	39,7
genügend, suffisant, Note 4	12,5
ungenügend, insuffisant, Note unter 4	1,3

In Entsprechung zur überwiegend positiven Benotung durch die Lehrpersonen nehmen die Schülerinnen und Schüler die Bewertung ihrer Arbeit mehrheitlich als gerecht wahr. Mehr als die Hälfte (52,4%) bezeichnet die eigene Note als voll angemessen, rund ein Drittel (32,7%) als eher angemessen. 11,1% bekunden, dass ihnen die Note eher nicht angemessen und 3,7%, dass ihnen die Note gar nicht angemessen erscheint. Dabei gibt es erwartungsgemäss einen starken Zusammenhang zwischen der Zufriedenheit mit der Note und der Note selbst. Die Rangkorrelation (Kendall-Tau b) fällt mit .52 hochsignifikant aus.

Von den 906 Personen, die auf die Frage geantwortet haben, ob ihnen die Bewertungskriterien vor Beginn der Maturaarbeit bekannt waren, gaben 82,9% an, dass sie die Kriterien kannten. 17,1% vermerkten hingegen, dass dies nicht der Fall war. Ob in diesen Fällen die Beurteilungskriterien nicht kommuniziert wurden, oder ob die Maturandinnen und Maturanden sich aus anderen Gründen nicht für ausreichend informiert hielten, ist unklar. Gemäss den schulischen Rahmenbedingungen gibt es in der Festlegung der Beurteilungskriterien unterschiedliche Vorgehen: Zwei Drittel der Schulen halten die Beurteilungskriterien schriftlich fest, knapp ein Drittel überschreibt die Definition und Aushandlung von Kriterien den betreuenden Lehrpersonen. Bei zwei Schulen (7%) finden sich keine Angaben zur Bewertung und Beurteilung. Eindeutig ist aber, dass mehr als drei Viertel der

nach eigenen Angaben nicht informierten Schülerinnen und Schüler die Kriterien gerne vorab gekannt hätten. Nur 22% der Nichtinformierten wäre dies nicht wichtig gewesen. Von den Schülerinnen und Schülern, denen die Kriterien vorab zur Information zur Verfügung standen, erachtet sie die grosse Mehrheit als eher sinnvoll (54,5%) oder sehr sinnvoll (39,6%). Nur insgesamt 5,9% meinen, dass die Kriterien nicht sinnvoll sind.

4.5 Betreuung

Die meisten Maturandinnen und Maturanden (59,1%) haben ihre Betreuungsperson für die Maturaarbeit aufgrund ihrer Themen oder Fächer wählen können, es gaben aber auch 28% der Befragten an, dass die Betreuungsperson von der Schule vorgegeben wurde, bei 20% der Befragten spielte ebenfalls die Sympathie eine Rolle (Mehrfachantworten waren möglich). Auffällig ist hierbei, dass in der französischsprachigen Schweiz wesentlich häufiger die Betreuung von der Schule vorgegeben wird und in der Deutschschweiz und im Tessin die Betreuungsperson häufiger aufgrund des Themas oder des Faches gewählt wird. Aus den institutionellen Vorgaben im Tessin geht jedoch auch hervor, dass die Entscheidung zugunsten eines bestimmten Projekts durch die Wahl eines Seminarrahmenthemas vorbestimmt wird. Eine Betreuungsperson ist jeweils im Rahmen der Seminarorganisation für ein Rahmenthema zuständig, so dass die Wahl der Betreuungsperson zwar über die Themenwahl geschieht, dennoch aber durch das jeweilige Seminarangebot der Lehrperson eingeschränkt ist. Der hohe Anteil der Tessiner Maturandinnen und Maturanden, die angeben, die Betreuungsperson aufgrund der Themen oder Fächer ausgewählt zu haben, ist durch diese vorgegebene Seminarstruktur zu erklären.

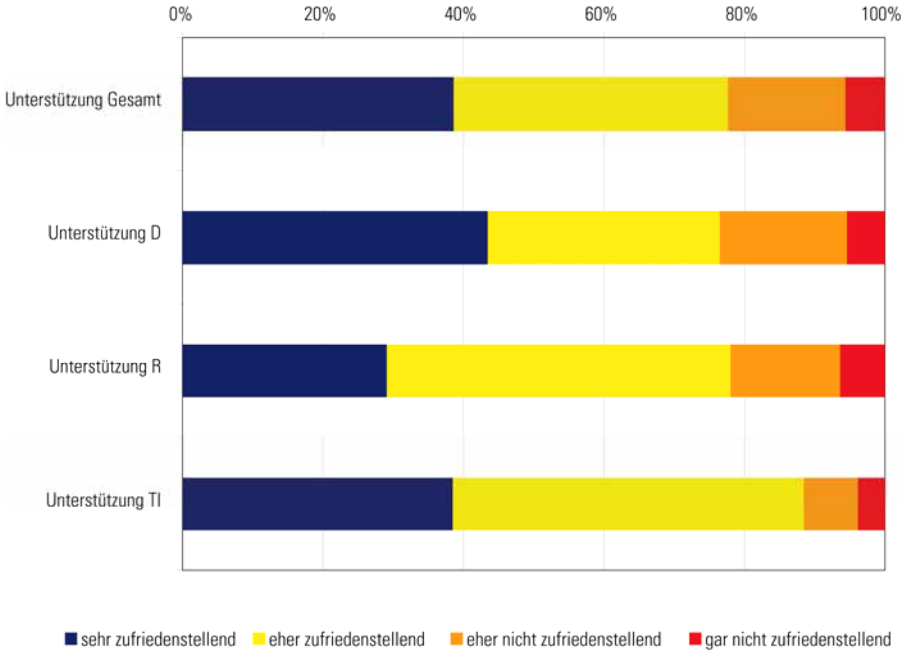
Tab. VI.15: Auswahl der Betreuungsperson nach Sprachregion. (Angaben in Prozent, Mehrfachantworten möglich)

Art der Themenwahl	Deutschschweiz	Romandie	Tessin	Gesamt
Wurde vorgegeben	11,2	64,1	7,3	28,2
Aufgrund Themen / Fächer	73,4	27,2	85,5	59,1
Aufgrund Sympathie	26,5	11,1	1,8	20,0

Die Betreuung der Maturaarbeit stützt sich auf unterschiedliche Formen der Kommunikation. Mehr als 90% der Schülerinnen und Schüler wurden in Form von individuellen Gesprächen betreut, rund ein Viertel nahm an Gruppengesprächen teil und die Hälfte der Befragten stand per E-Mail in Kontakt mit den Lehrpersonen.

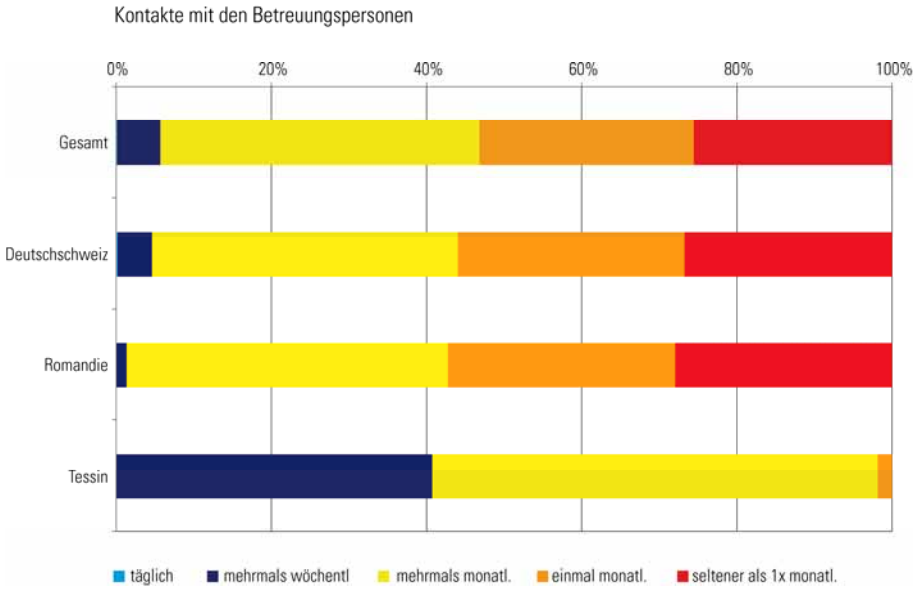
Insgesamt sind die Maturandinnen und Maturanden mit der Unterstützung durch die begleitende Lehrperson oder durch die begleitenden Lehrpersonen zufrieden. Abbildung VI.9 zeigt die Einschätzungen dazu nach Sprachregionen auf. Besonders die Schülerinnen und Schüler aus den zwei Tessiner Schulen fühlen sich durch ihre Lehrpersonen sehr gut unterstützt, aber auch in der Romandie, wo die Unterstützung vergleichsweise kritischer betrachtet wird, ist die grosse Mehrheit der Befragten mit der Unterstützungssituation zufrieden oder sehr zufrieden. Ähnliche Werte wurde auch bei der ersten EVAMAR-Studie erreicht, in deren Rahmen ebenfalls mehr als zwei Drittel der Befragten angaben, mit der Betreuung durch die Lehrpersonen zufrieden zu sein (Pagnossin et al., 2005, S. 187).

Abb. VI.9: Zufriedenheit mit der Unterstützung durch die begleitende(n) Lehrperson(en) (Angaben in Prozent)



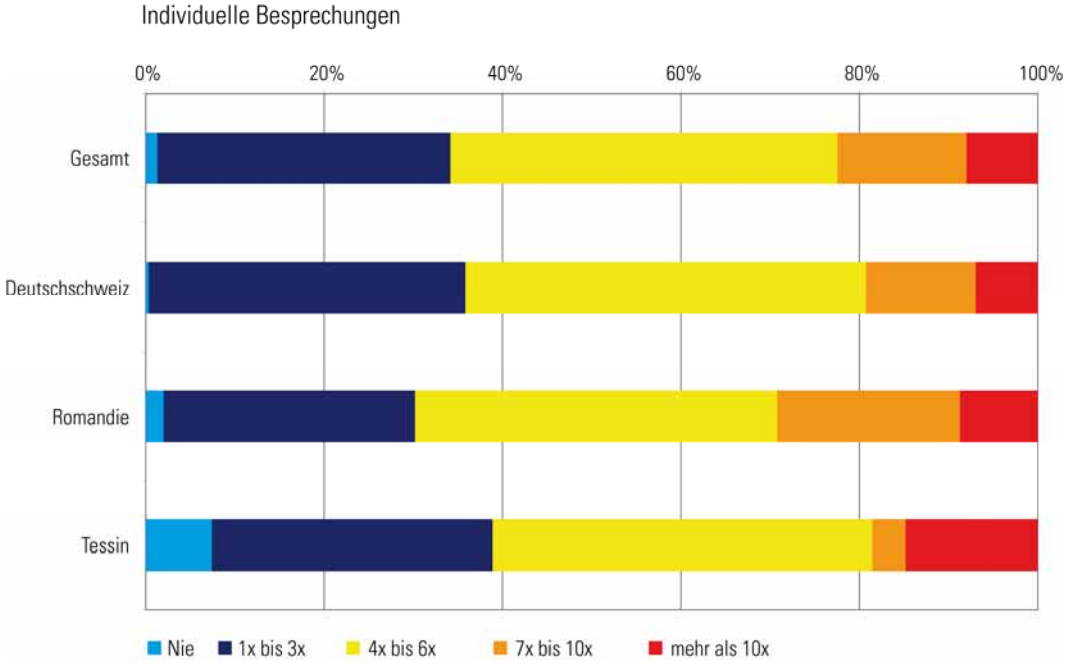
Über den Zeitraum der Durchführung des Projekts hatte ein Grossteil der Schülerinnen und Schüler mehrmals im Monat Kontakt mit ihren Betreuerinnen oder Betreuern. Jeweils ein Viertel wurde im Schnitt einmal im Monat oder weniger als einmal im Monat betreut. Beim Regionalvergleich (siehe Abbildung VI.10) fällt sofort die intensive Betreuung in den Tessiner Schulen ins Auge. Bedenkt man jedoch wiederum, dass die Maturaarbeiten in der italienischsprachigen Schweiz in eine Seminarstruktur eingebettet sind, ist dieses Ergebnis nicht mehr so erstaunlich. Die Aussagen der Maturandinnen und Maturanden in den anderen beiden Sprachregionen unterscheiden sich hingegen nur geringfügig. Auffällig ist in beiden Regionen, dass die Frequenz der pädagogischen Interaktion im Projektverlauf in vielen Fällen nicht sehr hoch ist. Ob dies mit einem mangelnden Angebot der Betreuungspersonen oder mit dem fehlenden Bedarf der Schülerinnen und Schüler zu erklären ist, können wir aus unseren Daten nicht direkt schliessen, wir kommen jedoch bei der Frage nach der Zufriedenheit mit der Betreuung darauf zurück.

Abb. VI.10: Häufigkeit der Kontakte mit den Betreuungspersonen (Angaben in Prozent)



Genaueren Aufschluss erlaubt auch die Auswertung der Frage nach der Häufigkeit der individuellen Besprechungen (siehe Abbildung VI.11). Ein Drittel der Befragten gibt an, dass während des Arbeitsprozesses ein bis drei solcher Gespräche stattgefunden haben, 43% hatten vier bis sechs solcher Gespräche. In mehr als einem Fünftel der Fälle ergab sich eine hohe oder sehr hohe Betreuungsfrequenz.

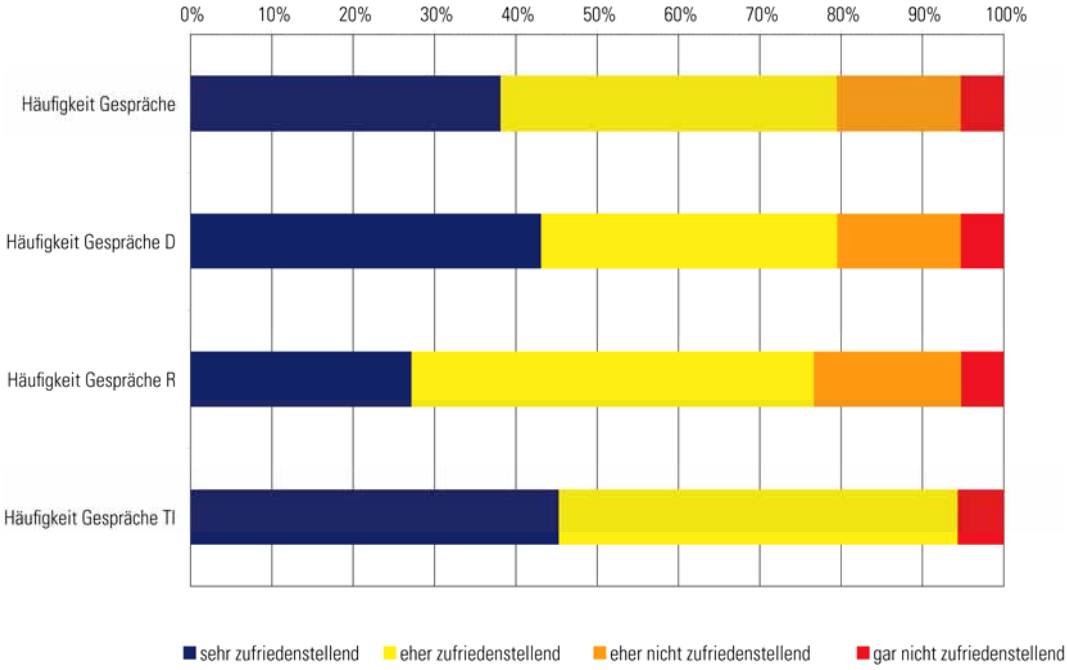
Abb. VI.11: Häufigkeit der individuellen Besprechungen mit den Betreuungspersonen (Angaben in Prozent)



Die Gelegenheit für individuelle Gespräche ist für viele Schülerinnen und Schüler relativ gering (siehe Abbildung VI.12). Wie oben bereits ausgeführt, kann von der Häufigkeit der Kontakte oder individuellen Gespräche nicht darauf geschlossen werden, ob das Angebot angemessen ist und ob der Bedarf für dieses Angebot besteht. Es muss zusätzlich auch betrachtet werden, ob die Häufigkeit der Gespräche von den Schülerinnen und Schülern als zufriedenstellend angesehen wird oder nicht.

Abb. VI.12 zeigt die Verteilungen zur Zufriedenheit mit der Häufigkeit der Gespräche. Insgesamt sind ca. 80% der Befragten mit der Häufigkeit der Gespräche eher zufrieden oder sehr zufrieden. Im Tessin ist dieser Anteil, vermutlich wieder wegen der Seminarorganisation, noch höher. Es besteht dabei ein starker Zusammenhang zwischen der Häufigkeit der Kontakte und individuellen Besprechungen und der Zufriedenheit mit der Häufigkeit der Gespräche. Im Rahmen der ersten EVAMAR -Studie konnte ein ähnliches, wenn auch etwas geringer ausfallendes Ergebnis festgestellt werden. 71% der damals befragten Schülerinnen und Schüler waren mit der Häufigkeit der Gespräche zufrieden (Pagnossin et al., 2005, S. 187).

Abb. VI.12: Zufriedenheit mit der Häufigkeit der Gespräche (Angaben in Prozent)



Nach der Art der bezogenen Beratung befragt (vgl. Tabelle VI.16), gaben jeweils etwa die Hälfte der Schülerinnen und Schüler an, die Hilfe der Lehrpersonen bei der Festlegung des Themas, bei inhaltlichen Fragen und bei Problemen hinsichtlich des Aufbaus und der Gliederung der Arbeit in Anspruch genommen zu haben. 37% der Befragten haben sich bei formalen Fragen beraten lassen und etwa ein Viertel hat Auskünfte zu Informationsquellen und Arbeitsmethoden erbeten. Rund ein Fünftel hatte bei der Zeitplanung und beim Schreibprozess Unterstützungsbedarf. Die Überwindung von Motivationsproblemen spielte offenbar nur eine untergeordnete Rolle (12%). Auch hinsichtlich der Nutzung von Computer und Internet waren nach eigener Einschätzung nur 6% auf Hilfe angewiesen.

Tab. VI.16: Unterstützung durch die Lehrpersonen (Mehrfachnennungen möglich)

Bereich in dem Unterstützung gesucht wurde	Prozent der Befragten
Themenwahl	16,7
Formulierung der Fragestellung	49,9
Zeitplanung und Arbeitsorganisation	21,5
Computer und Internetnutzung	6,0
Inhalt der Arbeit (fachlich)	53,3
Informationsquellen	27,4
Arbeitsmethoden	24,7
Verfassen der Arbeit (Schreibprozess)	20,1
Aufbau und Gliederung der Arbeit	51
Motivation / Überwindung von Krisen	12,1
Formale Grundlagen (z.B. Fussnoten, Bibliographie, Zitieren usw.)	37,1
Sonstiges	4,4

Grundsätzlich fällt auf, dass eine Vielzahl unterschiedlicher Beratungsleistungen bei den betreuenden Lehrpersonen nachgefragt wird. Unsicherheiten und entsprechender Beratungsbedarf scheinen dabei nicht in einer genau abgrenzbaren singulären Phase zu bestehen; vielmehr lassen sich grundsätzlich drei Meilensteine erkennen: erstens die Einstiegsberatung bei der Präzisierung der Fragen, zweitens die Klärung inhaltlicher Fragen, die sich während der Erarbeitungsphase stellen und drittens die Beantwortung von Fragen zur Gliederung und zur Textgestaltung in der Redaktionsphase. Lediglich die Beratung im Hinblick auf elektronische Quellen, die Hilfe bei Motivationskrisen sowie der Support bei der Themenwahl werden weniger in Anspruch genommen. Zumindest ersteres erstaunt kaum, wenn man davon ausgeht, dass die Schülerinnen und Schüler mit diesen Medien häufig vertrauter sind als ihre Lehrpersonen. Die wenig vorhandene Inanspruchnahme im Bezug auf die Motivation und die Themenwahl ist im Hinblick auf die Beweggründe der Wahl des Themas ebenfalls gut erklärbar, geben doch die Schülerinnen und Schüler mehrheitlich an, dass das Interesse am Thema die Hauptursache für die Wahl darstellt. Die Motivation scheint unter diesem Gesichtspunkt also kaum problematisch.

Die Frage nach der Art der in Anspruch genommenen Unterstützung lässt jedoch offen, aus welchen Gründen das Angebot angenommen oder nicht angenommen wurde, ob die Qualität der Beratung oder der Bedarf der Schülerinnen und Schüler dafür verantwortlich ist. Wir haben deshalb im Fragebogen gesondert nachgefragt, wie hilfreich die Schülerinnen und Schüler diese Unterstützung empfunden haben. Bei der Auswertung dieser Frage zeigt sich, dass fast alle Durchschnittswerte unterhalb des Skalenmittelpunkts von 2,5 liegen und damit von den Schülerinnen und Schülern als positiv bewertet werden. Im negativen Bereich liegt mit 2,77 allein das Item *Computer- und Internetnutzung*, was wenig erstaunt, da die Schülerinnen und Schüler, wie oben ausgeführt, häufig über gute Kenntnisse hinsichtlich der üblichen Anwenderprogramme verfügen. Die Unterstützung beim Verfassen der schriftlichen Arbeit und bei der Bewältigung von Motivationskrisen wird von den Schülerinnen und Schülern als vergleichsweise weniger hilfreich angesehen. Es ist anzunehmen, dass der Bedarf der Schülerinnen und Schüler in diesem Bereich auch wesentlich niedriger ist als in den anderen Bereichen, da die ex post deklarierte Nachfrage für eine Unterstützung im Schreibprozess oder bei Motivationskrisen ebenfalls geringer ist. Die Unterstützung bei der Formulierung der Fragestellung sowie beim Aufbau und bei der Gliederung der Arbeit wird von den Schülerinnen und Schülern am hilfreichsten wahrgenommen.

Tab. VI.17: Bewertung der Unterstützungsangebote (Mittelwerte, Standardabweichungen)

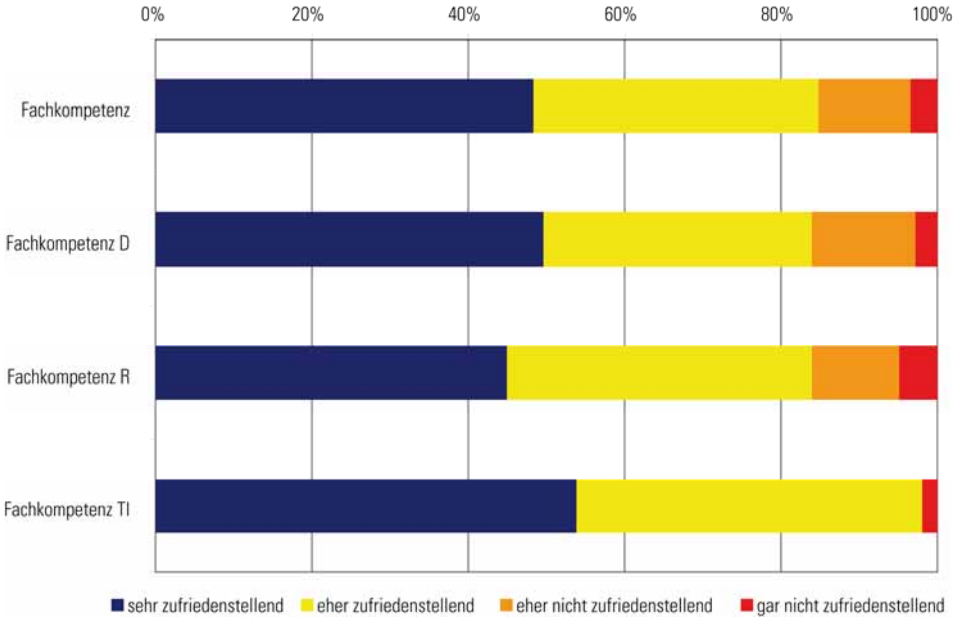
Bereich in dem Unterstützung gesucht wurde	Mittelwert	SD
Themenwahl	2,2	1,00
Formulierung der Fragestellung	1,9	0,82
Zeitplanung und Arbeitsorganisation	2,2	0,96
Computer und Internetnutzung	2,8	1,09
Inhalt der Arbeit (fachlich)	2,0	0,88
Arbeitsmethoden	2,2	0,93
Aufbau und Gliederung der Arbeit	1,9	0,84
Verfassen der Arbeit (Schreibprozess)	2,4	1,04
Motivation / Überwindung von Krisen	2,3	1,09
Formale Grundlagen (z.B. Fussnoten, Bibliographie, Zitieren usw.)	2,1	1,05
Sonstiges	1,7	1,07

1=sehr hilfreich, 2=eher hilfreich, 3=eher nicht hilfreich, 4=gar nicht hilfreich

Wichtig ist in diesem Zusammenhang auch die Einschätzung der Maturandinnen und Maturanden zur Fachkompetenz der betreuenden Lehrperson, da diese Einschätzung vermutlich einen Einfluss auf die Nachfrage der Unterstützung hat. Tatsächlich zeigt sich ein positiver Zusammenhang zwischen der geleisteten Unterstützung im inhaltlichen Bereich der Arbeit und der Einschätzung zur Fachkompetenz. (Spearman's Rho: .534). Zudem schätzen jene Schülerinnen und Schüler, die im fachlich-inhaltlichen Bereich pädagogische Unterstützung nachgefragt haben, die Fachkompetenz ihrer Betreuerinnen und Betreuer signifikant höher ein als jene, die nicht nach einer solchen Unterstützung verlangt haben (t-Test, zweiseitig, $\alpha<.005$).

Abbildung VI.13 zeigt die prozentualen Angaben zur Einschätzung der Fachkompetenzen. Auffällig ist hier die generell sehr hohe Einschätzung, die die Schülerinnen und Schüler von der Fachkompetenz ihrer Lehrpersonen haben, wobei die beiden Schulen im Tessin die besten Werte erzielen.

Abb. VI.13: Einschätzung der Fachkompetenzen (Angaben in Prozent)



Neben der Unterstützung durch die Lehrpersonen nennen viele Befragte noch weitere Personen, die ihnen bei der Abfassung der Arbeit behilflich gewesen sind. Knapp zwei Drittel verweisen auf die Eltern, knapp die Hälfte auf Freunde und Bekannte und 45% auf die Unterstützung durch Fachleute. Die Geschwister spielen hier mit 23% ebenso wie die Kolleginnen und Kollegen in der Schule mit knapp 18% eine deutlich geringere Rolle (Tabelle VI.18).[55]

Tab. VI.18: Weitere Unterstützung (Mehrfachantworten möglich)

Ausserschulische Unterstützung	Prozent der Befragten
Eltern	64,2
Geschwister	23,0
Freunde/Bekannte	47,8
Klassenkolleginnen/ -kollegen	17,9
Externe Fachpersonen	45,1

[55] Ein Grund für die geringe Bedeutung der Geschwister könnte darin liegen, dass nicht alle Schülerinnen und Schüler Geschwister haben. Da wir jedoch das Vorhandensein von Geschwistern nicht erhoben haben, bleibt unklar aus welchen Gründen die Geschwister auch gegenüber den Eltern eine so geringe Rolle spielen. Geht man entsprechend der amtlichen Statistiken davon aus, dass 3 von 10 Kindern ohne Geschwister aufwachsen, würde sich der Anteil derjenigen, die eine Unterstützung von vorhandenen Geschwistern erhalten haben, auf 32% erhöhen.

Maturandinnen und Maturanden, die eine Unterstützung aus einem ausserschulischen Netzwerk erhalten, greifen teilweise ebenfalls häufiger auf die Betreuung durch die Lehrpersonen zurück. Personen, die Unterstützung von Kolleginnen und Kollegen erhalten haben, haben häufiger Kontakt zu den begleitenden Lehrpersonen (t-Test, zweiseitig, $\alpha<0{,}05$) und Personen, die sich zusätzlich an externe Fachpersonen gewandt haben, haben häufiger individuelle Besprechungen mit ihren Betreuungspersonen (t-Test, zweiseitig, $\alpha<0{,}05$). Die Inanspruchnahme anderer Unterstützung, wie z.B. durch die Eltern, steht nicht in Zusammenhang mit der Betreuung durch die Lehrpersonen. Auffällig daran ist, dass es sich bei den Kontakten mit Kolleginnen und Kollegen sowie mit externen Fachpersonen eher um Kontakte zu handeln scheint, die aus einem fachlichen oder inhaltlichen Interesse heraus geknüpft werden. Insoweit verwundert es nur wenig, dass in diesen Fällen auch eine engere Zusammenarbeit mit der betreuenden Lehrperson gesucht wird. Im Fall des Austausches mit externen Fachpersonen ist weiterhin denkbar, dass die Lehrperson selbst bei der Kontaktaufnahme mit dieser Person eine Rolle spielt und daher auch eine intensivere Unterstützung geleistet hat. Die Daten zeigen deutlich, dass die Schülerinnen und Schüler die externen Fachpersonen nicht als Ersatz für ihre Betreuungspersonen ansehen, sondern als Ergänzung zu deren Fachkompetenz. Insofern ist der hohe Anteil der Maturandinnen und Maturanden, die bei externen Fachpersonen Unterstützung gefunden haben, eher als ein Zeichen für eine gute Kooperation zwischen Schülerinnen und Schülern und den Lehrpersonen anzusehen.

Dass die Eltern, Freunde und Bekannte beim Verfassen der Maturaarbeit für viele Maturandinnen und Maturanden eine so grosse Rolle spielen, ist insofern ein positiver Befund, da sich offensichtlich die Schülerinnen und Schüler bewusst sind, dass bei grösseren Projekten der intensive Austausch mit anderen Personen hilfreich ist. Für die Maturandinnen und Maturanden bietet es sich an, hier neben den Lehrkräften vor allem auf den Familien- und Bekanntenkreis zurückzugreifen. Problematisch ist das insofern, weil es sich hier um soziales Kapital handelt, das in der Bevölkerung nicht gleich verteilt ist. Hier können sich mithin soziale Ungerechtigkeiten ergeben, wenn die zur Verfügung stehenden ideellen und materiellen Unterstützungsmöglichkeiten von Familie zu Familie sehr stark variieren. Ob solche Ungerechtigkeiten durch die betreuenden Lehrerinnen und Lehrer vollständig ausgeglichen werden können, ist fraglich.

4.6 Institutionelle Rahmenbedingungen und individuelle Voraussetzungen

Ungefähr drei Viertel der befragten Schülerinnen und Schüler (77,3%) bejahen, dass an ihrer Schule ein Leitfaden für das Verfassen der Maturaarbeit existiert und 93,3% von ihnen geben an, dass ihnen dieser auch während des Arbeitsprozesses bekannt gewesen war. Umgekehrt wussten 22,7% der Befragten nichts über einen Leitfaden, oder bestritten ausdrücklich, dass es an ihrer Schulen einen solchen gäbe. Dass diese Schülerinnen und Schüler aber trotzdem an einer solchen Orientierungshilfe interessiert gewesen wären, zeigt der sehr hohe Anteil von 66,9% dieser Befragten, die angeben, dass ihnen ein Leitfaden eher wichtig oder sehr wichtig gewesen wäre. Ein solcher Befund ist insofern bemerkenswert, da gemäss Analyse der Rahmenbedingungen sämtliche Schulen über einen Leitfaden verfügen (vgl. Kapitel 3). Einerseits scheint also die grosse Mehrheit der Schulen ihre Leitfäden tatsächlich als Hilfsdokument einzusetzen, andererseits gibt ca. ein Viertel der Schülerinnen und Schüler an, diese schuleigenen Richtlinien nicht zu kennen. Geht es um den Gebrauchswert dieser Richtlinien für die Lernenden, so ergibt sich im Hinblick auf die Implementation von Leitfäden für viele Kantonsschulen ein Handlungsbedarf - aber nicht für alle: In diesem Punkt sind die Unterschiede zwischen den Gymnasien beträchtlich. Geht man davon aus, dass der Grund für die fehlende Kenntnis nicht bei den Schülerinnen und Schülern, sondern vorrangig bei den Gymnasien selbst zu suchen ist, sollte sich der Einfluss der Schulzugehörigkeit statistisch bemerkbar machen. Tatsächlich kann in diesem Fall 38,6% der Varianz durch die Schulzugehörigkeit erklärt werden. Während in einigen Schulen alle Schülerinnen und Schüler angeben, die Leitfäden zu kennen, beschränkt sich diese Kenntnis andernorts nur auf eine Minderheit.

Der Nutzen der Leitfäden wird von 61,2% der Befragten, die Kenntnis von ihnen haben, als eher hoch oder sehr hoch eingeschätzt, rund ein Drittel kommt jedoch zu einer eher negativen Bewertung. Selbstverständlich gibt es bei der Einschätzung zur Nützlichkeit der Leitfäden ebenfalls grosse Unterschiede zwischen den Schulen, da die

Schülereinschätzungen sich ja auf unterschiedliche Leitfäden beziehen. 16,7% der Varianz bei dieser Einschätzung geht auf die Schulzugehörigkeit zurück.

In vielen Schulen werden vorbereitende Veranstaltungen für die Bearbeitung der Maturaarbeit angeboten. Mehr als die Hälfte der Schülerinnen und Schüler hat eine solche Veranstaltung besucht, 42,4% geben jedoch an, dass an ihrer Schule keine vorbereitende Veranstaltung stattgefunden hat, obwohl 35,2% von ihnen diese für eher wichtig oder sehr wichtig gehalten hätte.

Tab. VI.19: Vorbereitende Veranstaltung (Mehrfachnennungen)

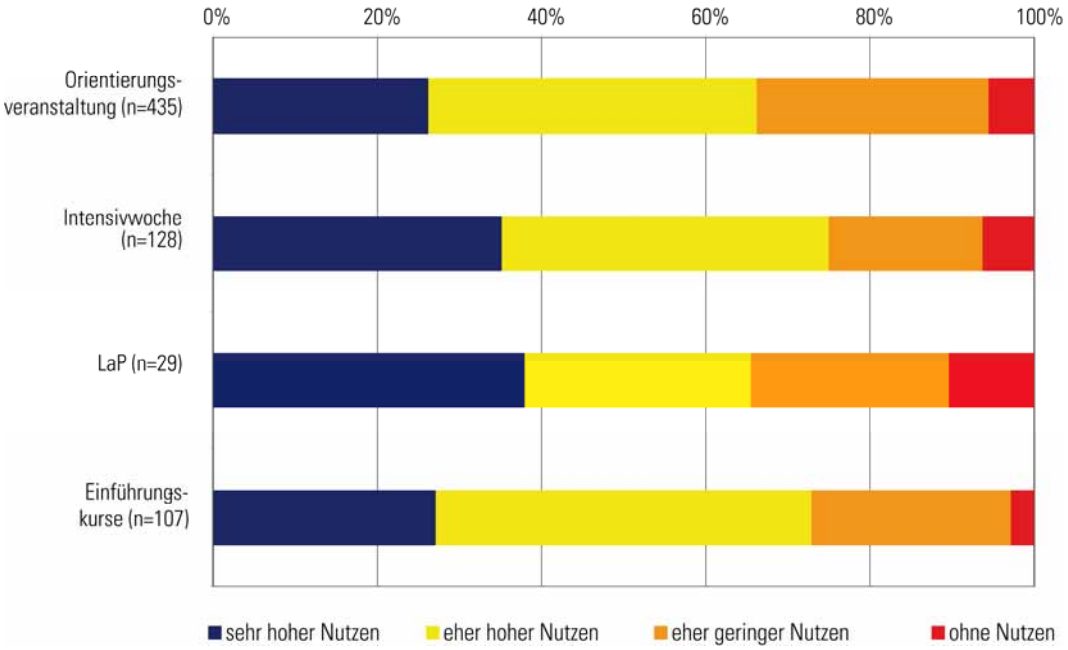
Veranstaltungsart	Prozent der Befragten
Orientierungs- / Informationsveranstaltung	82,7
Intensivwoche / Projektwoche	24,7
Lernen am Projekt LaP	6,0
Einführungskurse	21,7
Andere	7,0

Wie aus Tabelle VI.19 hervorgeht, handelt es sich bei den besuchten Veranstaltungen mehrheitlich um Orientierungs- und Informationsveranstaltungen, etwa ein Fünftel der Maturandinnen und Maturanden hat einen Einführungskurs besucht, ein Viertel eine Intensiv- oder Projektwoche.[56] Der Nutzen dieser Veranstaltungen wird im Durchschnitt als eher hoch bewertet. Allerdings bekunden auch jeweils ca. ein Drittel der Maturandinnen und Maturanden, die eine entsprechende Veranstaltung besucht haben, dass diese ohne oder nur von geringem Nutzen sei.

Erwartungsgemäss gibt es in diesem Zusammenhang schulspezifische Unterschiede. Die Einschätzung des Nutzens von Orientierungs- und Informationsveranstaltungen und Intensiv- oder Projektwochen fällt je nach Schulzugehörigkeit verschieden aus, wobei die dadurch erklärte Varianz bei ca. 17% liegt. In Abbildung VI.14 können die Einschätzungen zum Nutzen der Veranstaltungen im Überblick eingesehen werden.

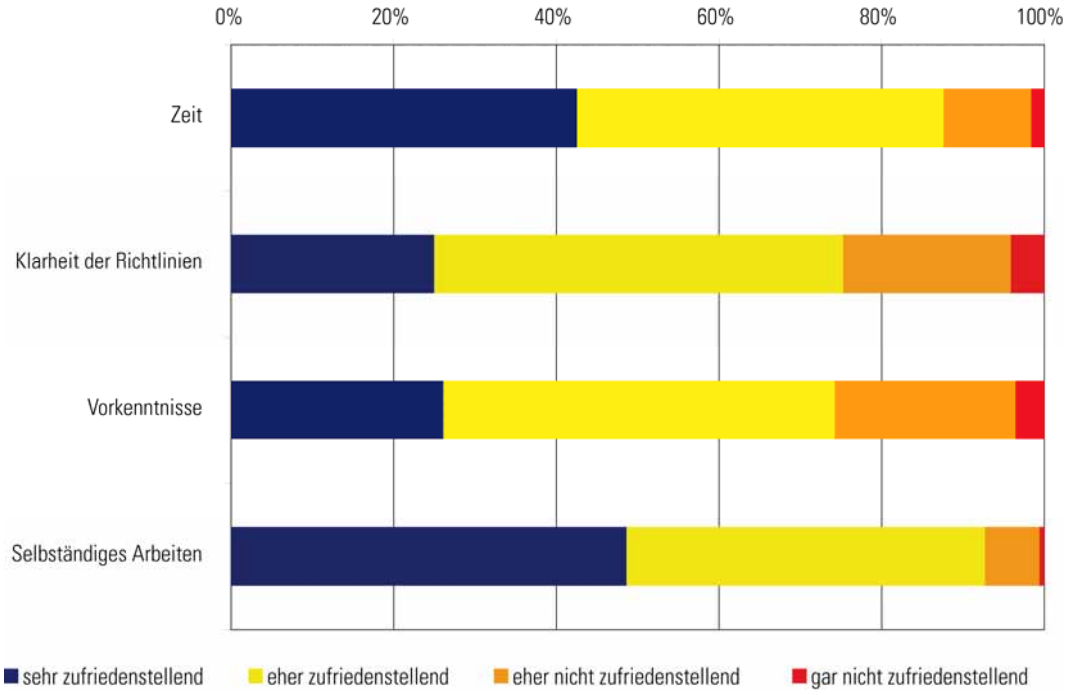
[56] Nicht miteinbezogen sind die vorbereitenden Elemente in den spezifischen Lehrveranstaltungen, innerhalb derer in den meisten Schulen der lateinischen Schweiz die Maturaarbeiten geplant und durchgeführt werden (vgl. Kapitel 3).

Abb. VI.14: Einschätzungen von Teilnehmerinnen und Teilnehmern zum Nutzen vorbereitender Veranstaltungen (Angaben in Prozent)



Im Rahmen der Untersuchung zu den Realisierungsbedingungen für die Maturaarbeit haben wir auch die Zufriedenheit mit den Rahmenbedingungen erfragt. Dabei spielt die Zeit eine Rolle, die den Maturandinnen und Maturanden zur Verfügung stand, die Klarheit der Richtlinien, die Vorkenntnisse, um eine selbstständige Arbeit realisieren zu können und die Fähigkeit, selbstständig zu arbeiten. Wie Abbildung VI.15 deutlich macht, liegen die Einschätzungen der Schülerinnen und Schüler hinsichtlich dieser Determinanten zum grossen Teil im positiven Bereich. Während die Maturandinnen und Maturanden in der ersten EVAMAR-Studie zu 71,9% mit den zeitlichen Ressourcen und zu 55,6% mit der Klarheit der Richtlinien zufrieden waren (Pagnossin et al., 2005, S. 187), sind in dieser Befragung 87% damit zufrieden, wie viel Zeit ihnen für die Bearbeitung ihrer Maturaarbeit zur Verfügung gestellt wurde und 75,3% sind mit der Klarheit der Richtlinien zufrieden. Die Zusammensetzung der Stichproben scheint bei dieser grossen Differenz der Einschätzungen keinen Einfluss zu haben, da unsere Ergebnisse sich auch mit einer Gewichtung nach Sprachregion und Geschlecht nicht wesentlich verändern. Eventuell spielt aber der Zeitpunkt der Befragung dabei eine Rolle. Da in unserer Untersuchung die Maturaarbeit für die Befragten soweit zurücklag, dass sie bereits Kenntnis über die Ergebnisse hatten, könnte es sein, dass sie im Nachhinein die Rahmenbedingungen ihrer Arbeit in ein etwas günstigeres Licht gestellt haben. Ebenfalls ist aber auch denkbar und zu hoffen, dass die Erfahrung, die die Schulen in der Zwischenzeit mit der Organisation der Maturaarbeiten gemacht haben, sich positiv auf die Rahmenbedingungen und darüber auf die Einschätzungen der Schülerinnen und Schüler ausgewirkt haben.

Abb. VI.15: Institutionelle und individuelle Realisierungsbedingungen für die Maturaarbeit



Mit ihren eigenen Fähigkeiten zum selbstständigen Arbeiten sind die Schülerinnen und Schüler fast durchgehend zufrieden, es gibt nur sehr wenige Befragte, die hier mit "eher nicht zufriedenstellend" oder "gar nicht zufriedenstellend" geantwortet haben. Eher sind sie kritisch gegenüber ihren Vorkenntnissen eingestellt, aber auch dies trifft nur für ca. ein Viertel der Befragten zu. Die Ergebnisse sind mit denen aus EVAMAR I vergleichbar. Dort waren 68,7% mit ihren Vorkenntnissen und 91,1% mit ihrer Fähigkeit zum selbstständigen Arbeiten zufrieden (Pagnossin et al., 2005, S. 187).

4.7 Einschätzungen zu Aufwand und Nutzen der Maturaarbeit

Hinsichtlich der tatsächlich aufgewendeten Zeit differieren die Angaben der Schülerinnen und Schüler beträchtlich. Einerseits gibt ein Viertel der Befragten an, mehr als 150 Stunden für die Maturaarbeit aufgewendet zu haben, andererseits liegt bei rund 28% der wahrgenommene Arbeitsaufwand bei weniger als 80 Stunden.[57] Der Median des Aufwands liegt im Intervall zwischen 100 und 120 Stunden bei 106,6. Die grosse Diskrepanz zwischen den für die Maturaarbeiten eingesetzten Arbeitsstunden lässt sich wohl nicht einfach nur auf individuelle Unterschiede im Arbeitstempo zurückführen; hier liegt die Vermutung nahe, dass beim Zeitaufwand weitere strukturelle Unterschiede zwischen Schulen oder auch Fachbereichen zum Tragen kommen. Bei Arbeiten aus dem sprachlichen Bereich liegt die Bearbeitungszeit in der Regel etwas höher (Median=113,7 Stunden) und bei Arbeiten im sozialwissenschaftlichen Bereich tendenziell niedriger (Median=98,8). Hinsichtlich der Sprachregionen gibt es keine wesentlichen Unterschiede. Es muss bei der Interpretation der Daten allerdings auch bedacht werden, dass es sich um nachträgliche Einschätzungen der Schülerinnen und Schüler handelt und nicht um dokumentierte Daten. Es ist durchaus möglich, dass einzelne Befragte bei der Selbstdeklaration dazu neigen, die tatsächliche Arbeitszeit zu unterschätzen, während andere den Aufwand im Nachhinein überschätzen.

[57] Für die Vereinheitlichung der zu investierenden Arbeitszeit bieten die Kantonsschulen selbst kaum Fingerzeige: Nur einige wenige Schulen beziffern den zu erwartenden Arbeitsaufwand in ihren Richtlinien (vgl. Kapitel 3).

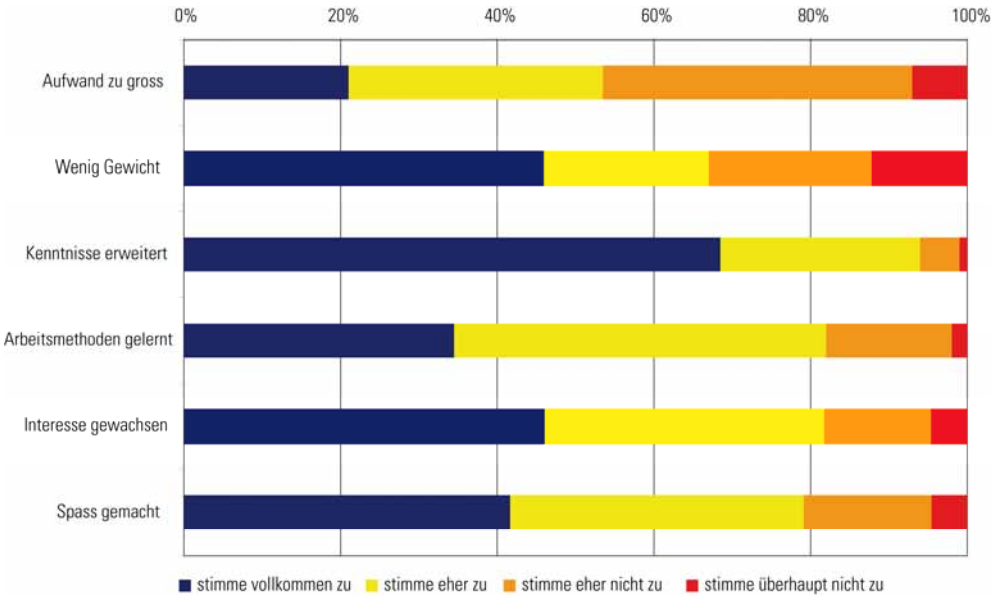
Tab. VI.20: Einschätzungen zum tatsächlichen individuellen Zeitaufwand

Geschätzter Zeitaufwand in Stunden	Prozent der Befragten
Weniger als 20	0,5
20-50	10,5
50-80	17,4
80-100	17,3
100-120	13,0
120-150	14,8
Mehr als 150	26,5

Viele der Maturandinnen und Maturanden hatten die Gelegenheit, sich auch im Rahmen regulärer Unterrichtsstunden mit ihrer Maturaarbeit zu beschäftigen. Im Durchschnitt konnten 16,3 Unterrichtsstunden für die Maturaarbeit genutzt werden. Hier ist allerdings die Spannweite beträchtlich: Sie reicht von 0 bis 100 Stunden, wobei 27,9% der Befragten angaben, dass überhaupt keine Unterrichtszeit für die Maturaarbeit zur Verfügung stand und weitere 26,8% auf diese Frage nicht antworteten. Der Umstand, dass ca. 50% der Maturandinnen und Maturanden nach eigener Auskunft keine Bearbeitungszeit während der Unterrichtsstunden zur Verfügung stand, deckt sich im Übrigen mit den Ergebnissen der von uns durchgeführten Analysen zu den Rahmenvorgaben der Schulen.

Die für die Arbeit aufgewendete Zeit wird von vielen Befragten im Vergleich zu den anderen Anforderungen der Matura als zu lang angesehen. 53,5% der Befragten stimmen dieser Aussage eher oder vollkommen zu. Dies mag zu grossen Teilen damit zusammenhängen, dass die Arbeit für das Erlangen des Maturitätszeugnisses wenig ins Gewicht fällt, was 67% der Befragten bemängeln. Neben der Einschätzung zum grossen Zeitaufwand sind die Maturandinnen und Maturanden aber zu grossen Teilen auch davon überzeugt, dass sie durch das Verfassen der Maturaarbeit viel gelernt haben. 94% von ihnen geben an, dass sie ihre Kenntnisse erweitern konnten, 82% haben viel über Arbeitsmethoden gelernt, bei 81,8% konnte wiederum ein erweitertes Interesse für das Thema geweckt werden. Weitere 79,1% geben an, dass ihnen die Arbeit Spass gemacht hat. Abbildung VI.16 stellt diese Einschätzungen zu Aufwand und Nutzen dar.

Abb. VI.16: Einschätzungen zu Aufwand und Nutzen (Angaben in Prozent)

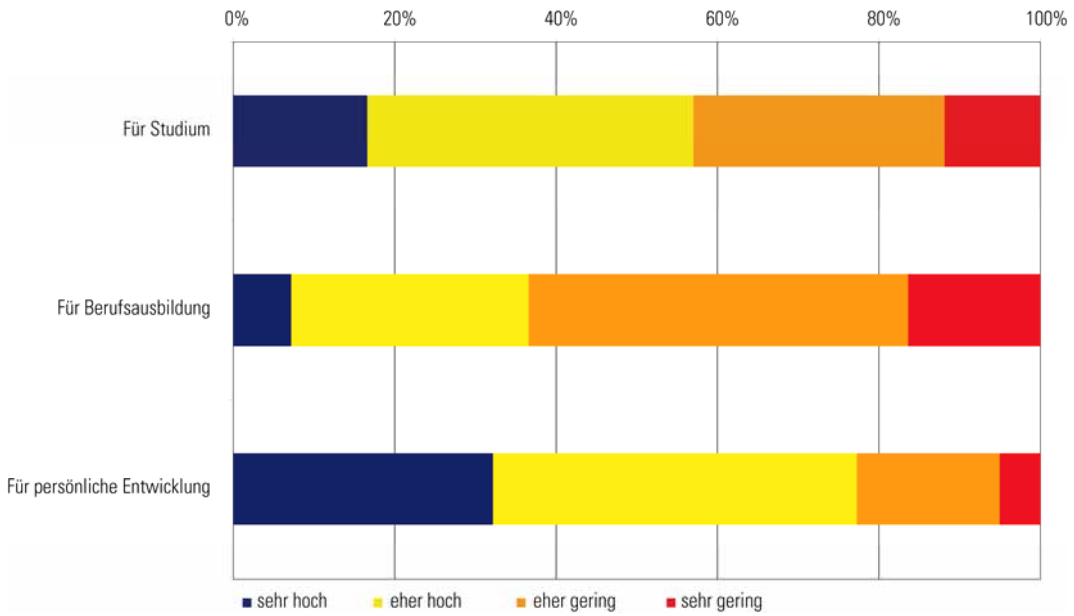


Hinsichtlich des Geschlechts gibt es keine Unterschiede in den Einschätzungen zum Aufwand, vergleicht man jedoch die Mittelwerte zu den einzelnen Einschätzungen nach Sprachregionen so fallen in drei der sechs Einschätzungen signifikante Gruppenunterschiede auf. Die Maturandinnen und Maturanden in der französischsprachigen Schweiz sind gegenüber der Aussage, sie hätten ihre Kenntnisse erweitert, eher skeptisch. Dagegen glauben sie aber eher, dass die Maturaarbeit im Maturitätszeugnis ein angemessenes Gewicht hat. Den Tessiner Maturandinnen und Maturanden hat das Verfassen der Arbeit grundsätzlich eher weniger Spass gemacht, als ihren Kolleginnen und Kollegen aus den anderen Sprachregionen.[58]

In mancherlei Hinsicht lassen sich unsere Ergebnisse als Bestätigung von Indizien aus der ersten EVAMAR-Studie interpretieren. Im Rahmen der Erhebungen von EVAMAR I hatte sich ergeben, dass rund drei Viertel der Maturandinnen und Maturanden ein positives Urteil über die Projektarbeit abgeben: 88% bezeugten einen Kenntniszuwachs, 78% bezeichneten die Realisierungsbedingungen als gut und 79% sagten, dass ihr Interesse am Thema während des Projekts gewachsen sei. Gemessen am geleisteten Aufwand, schätzten 62% der damals befragten Absolventinnen und Absolventen das Gewicht der Maturaarbeit für die Erlangung des Maturitätszeugnisses als zu gering ein; ähnlich hoch war der Wert für die kritische Aussage, dass der zu leistende Aufwand für die Maturaarbeit im Kontext der Maturitätsprüfungen zu hoch gewesen sei (Pagnossin et al., 2005, S. 189). Starke regionale Unterschiede traten dabei nicht zu Tage, allerdings fiel auf, dass in der Romandie das fehlende Gewicht für das Zertifikat im Vergleich zu den anderen Landesteilen weniger stark beklagt wurde. Während in der Deutschschweiz und in der Romandie rund zwei Drittel den Aufwand für ihr jeweiliges Projekt als zu gross erachteten, lag dieser Wert im Tessin knapp unter 50% (Pagnossin et al., 2005, S. 190).

Hinsichtlich der Auswirkungen der Maturaarbeit auf den weiteren Qualifikationsprozess ergibt sich in unserer Untersuchung in Bezug auf das Studium eine relativ hohe Nutzeneinschätzung der Maturandinnen und Maturanden. 57% der Befragten schätzen den Nutzen eher hoch oder sogar sehr hoch ein. Der Nutzen für die Berufsausbildung wird als vergleichsweise gering angesehen. Hier sind es rund 38%, die den Nutzen hoch bis sehr hoch einschätzen. Am stärksten fällt der Nutzen für die persönliche Entwicklung ins Gewicht: Mehr als drei Viertel der Befragten geben hier eine positive oder sehr positive Einschätzung ab (siehe Abbildung VI.17).

Abb. VI.17: Einschätzungen im Hinblick auf Qualifikation und persönliche Entwicklung



[58] Dabei ist nicht auszuschliessen, dass hier die Wortwahl eine wichtige Rolle spielt: Während im Deutschen von "Spass" die Rede ist, ergibt sich im Französischen bei diesem Item eine Konnotation zu "Wertschätzung" (appréciation) und im Italienischen zu "Ablenkung" oder "Abwechslung" (divertimento).

Um genauere Angaben zu der Wirkung der Maturaarbeit zu bekommen, wurden die Maturandinnen zu ihrer Einschätzung befragt, wie sich die Maturaarbeit bei ihnen auf einzelne Kompetenzaspekte ausgewirkt hat. Betrachtet man die Mittelwerte in Tabelle VI.21, so zeigt sich, dass die Maturaarbeit aus Sicht der Befragten keine negativen Auswirkungen hatte - womit indes noch kein pädagogisches Ziel erreicht wäre. Unverkennbar gibt es einzelne Aspekte, die aus Sicht der Akteure im Verlauf der Projekte wenig gefördert wurden. Dazu gehören vor allem die Teamfähigkeit, das Abstraktionsvermögen, das logische Denken, die Vorstellungskraft und das vernetzte Denken. Es liegt auf der Hand, dass Einzelarbeiten für die Teamfähigkeit weniger förderlich sind als Gruppenarbeiten. Tatsächlich liegt der Mittelwert zur Teamfähigkeit bei den 54 Personen, die ihre Arbeit im Team erstellt haben bei 1,8 gegenüber 2,7 bei Einzelarbeiten. Auch das Abstraktionsvermögen, das logische Denken die Vorstellungskraft und das vernetzte Denken sind je nach Themenwahl unterschiedlich gefordert, Allerdings lässt sich die Vermutung, dass dies vor allem im mathematisch-naturwissenschaftlichen Bereich der Fall ist, nicht erhärten. Es gibt kaum Unterschiede in der Verteilung, wenn ausschliesslich die in einem naturwissenschaftlichen Hauptfach geschriebenen Maturaarbeiten betrachtet werden.

Tab. VI.21: Wirkungen der Maturaarbeit

Wie hat sich das Verfassen der Maturaarbeit bei Ihnen auf die folgenden Aspekte ausgewirkt?	Mittelwert	SD
Teamfähigkeit	2,7	0,84
Selbstdisziplin	1,9	0,72
Selbstständiges Recherchieren	1,8	0,69
Selbstständiges Urteilen	1,9	0,70
Schriftliche Kommunikationsfähigkeit	2,1	0,75
Mündliche Kommunikationsfähigkeit	2,2	0,77
Vorstellungskraft	2,4	0,76
Logisches Denken	2,5	0,73
Abstraktionsvermögen	2,6	0,68
Vernetztes Denken	2,3	0,76
Persönliche Reife	1,9	0,77

Legende: 1=sehr positiv, 2=eher positiv, 3=keine Auswirkung, 4=eher negativ, 5=sehr negativ

Um die Einschätzungen zu den Wirkungen der Maturaarbeit übersichtlicher darzustellen und weiterzuverarbeiten, wurden Skalen zu den spezifischen Bereichen gebildet, in denen die Beschäftigung mit der Maturaarbeit nach Angabe der Maturandinnen und Maturanden eine Wirkung gezeigt hat. Dazu wurde eine explorative Faktorenanalyse durchgeführt, die aufzeigt, dass die Entwicklungsaspekte sich zu drei Hintergrundvariablen verdichten lassen. Der erste Faktor umfasst Aspekte der selbstständigen Arbeitsorganisation. Dazu gehören das selbstständige Recherchieren, die Selbstdisziplin, das selbstständige Urteilen, die schriftliche Kommunikation und die persönliche Reife. Der zweite Faktor setzt sich aus Aspekten der kognitiven Kompetenz zusammen. Dazu sind das Abstraktionsvermögen, das logische Denken, die Vorstellungskraft, und das vernetzte Denken zu zählen. Der dritte Faktor Aspekte der sozialen Kompetenz setzt sich nur aus den zwei Variablen Teamfähigkeit und mündliche Kommunikation zusammen. Insgesamt können mit diesen drei Faktoren 57,7% der Varianz abgebildet werden, wobei auf den ersten nach Varimax-Methode rotierten Faktor 23,2% entfallen, auf den zweiten Faktor 22,7% und auf den dritten Faktor weitere 12,2%.

Die Variablen der ersten beiden Faktoren wurden zu Summenskalen zusammengefasst, um sie für weitere Analysen bereitzustellen. Dabei erreichen beide Skalen eine ausreichend hohe interne Konsistenz. Für beide Skalen liegt Cronbach's Alpha bei .75. Für den Faktor *Soziale Kompetenz* wird auf eine Skalenbildung verzichtet, da es sich nur um zwei Items handelt.

Tab. VI.22: Skalen "Arbeitsorganisation" und "Kognitive Kompetenz"

"Wie hat sich das Verfassen der Maturaarbeit auf die folgenden Aspekte ausgewirkt?"

	Item	Kennwerte	Antwortmöglichkeiten
Arbeitsorganisation	Selbstständiges Recherchieren	gültige Fälle: n=902 Cronbach's Alpha: .755 M=1,9 SD=.51	sehr positiv eher positiv keine Auswirkung eher negativ sehr negativ
	Selbstdisziplin		
	Selbstständiges Urteilen		
	Schriftliche Kommunikation		
	Persönliche Reife		
	Item	**Kennwerte**	
Kognitive Kompetenz	Abstraktionsvermögen	gültige Fälle: n=894 Cronbach's Alpha: .751 M=2,49 SD=.56	
	Logisches Denken		
	Vorstellungskraft		
	Vernetztes Denken		

Für die Skala *Selbstständige Arbeitsorganisation* ergibt sich ein Gesamtmittelwert von 1,9 (SD=0,51), der deutlich im positiven Bereich liegt. Die Maturandinnen und Maturanden haben also im Schnitt den Eindruck, in diesem Bereich relativ stark profitiert zu haben. Der Gesamtmittelwert der Skala *Kognitive Kompetenz* fällt mit 2,4 (SD=0,56) im Vergleich dazu weniger positiv aus. Im Regionalvergleich zeigen sich keine Unterschiede. Lediglich die Tessiner Maturandinnen und Maturanden geben an, einen etwas grösseren Nutzen im Bereich der kognitiven Kompetenz gezogen zu haben. Die Fächer, in denen die Maturaarbeit geschrieben wurde, scheinen keine Rolle dabei zu spielen. Interessant ist jedoch ein beobachteter negativer Zusammenhang beider Skalenwerte mit der investierten Zeit. Je länger also die Maturandinnen und Maturanden sich nach eigener Erinnerung mit der Arbeit auseinandergesetzt haben, desto geringer ist deren zugeschriebene Wirkung auf die kognitive Kompetenz oder die selbstständige Arbeitsorganisation.

Weiterhin ist auffällig, dass die Einschätzung des Nutzens der Maturaarbeit für ein zukünftiges Studium, eine Berufsausbildung oder für die persönliche Entwicklung in engem Zusammenhang mit der Einschätzung zur Wirkung auf die kognitive Kompetenz und die selbstständige Arbeitsorganisation steht. Tabelle VI.23 stellt diese Zusammenhänge im Überblick dar.

Tab. VI.23: Korrelation von Nutzen- und Kompetenzaspekten (selbstattribuiert)

Nutzen	Kognitive Kompetenz	Arbeitsorganisation
Für Studium	0,9**	0,24**
Für Berufsausbildung	0,18**	0,26**
Für pers. Entwicklung	0,32**	0,46**

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Es scheint demnach bei den Schülerinnen und Schülern ein Bewusstsein für den wissenschaftspropädeutischen Nutzen der Maturaarbeit zu bestehen. Diejenigen Schülerinnen und Schüler, deren kognitiven Kompetenzen und Fähigkeiten zur selbstständigen Arbeitsorganisation nach eigener Wahrnehmung durch die Maturaarbeit geförderten wurden, sehen diese in der Regel auch als nützlich für ihre zukünftige persönliche oder berufliche Entwicklung an.

4.8 Zwischenbilanz

Die Analysen der Daten zur Befragung der Maturandinnen und Maturanden machen deutlich, dass zwar bei der Mehrheit der Befragten die Betreuung nicht sonderlich intensiv war, aber dennoch zu weiten Teilen als sehr zufriedenstellend empfunden wurde. Auch wenn es sich hier um eine retrospektive Einschätzung aus der Akteursperspektive handelt, bei der der Halo-Effekt der überwiegend positiven Benotung mit ins Spiel kommt, lässt das den Schluss zu, dass der Bedarf an Betreuung im Regelfall durch die Institutionen gut abgedeckt wird. Auch mit den Produkten selbst sind die Befragten mehrheitlich zufrieden. Herauszuheben ist dabei, dass sie vielfach angeben, ihre Kenntnisse erweitert und viel über Arbeitsmethoden gelernt zu haben. Zudem bekunden die meisten, dass ihr Interesse im Projektverlauf zugenommen und die Arbeit selbst Spass gemacht habe. Dementsprechend wird der Nutzen des Projekts für die persönliche Entwicklung von den meisten Maturandinnen und Maturanden hoch eingestuft. Die Einschätzungen im Hinblick auf den Ausbau qualifikationsrelevanter Fähigkeiten stimmen damit nicht überein: Die Nutzenerwartung im Hinblick auf akademische oder berufliche Kompetenzen fällt deutlich geringer aus.

Im Vergleich zu den Ergebnissen aus EVAMAR I ist zu registrieren, dass die Zufriedenheit der Maturandinnen und Maturanden in der hier vorliegenden Studie durchweg leicht höher ausfällt. Besonders bei der Beurteilung der schulischen Rahmenbedingungen, wie etwa in Hinblick auf die Transparenz der Richtlinien oder das Angebot für zusätzliche Veranstaltungen, sind die Unterschiede deutlich erkennbar. Dies kann zum einen dadurch erklärt werden, dass die Schülerinnen und Schüler unserer Studie zum Zeitpunkt der Erhebung bereits Kenntnis über ihre Bewertung hatten und deshalb ihre Bewertung im Rückblick weniger kritisch ausgefallen sein mag. Zum anderen ist die Erfahrung, die in den Schulen seit der Durchführung der ersten EVAMAR-Studie gesammelt wurde, als ein weiterer wichtiger Grund anzugeben. Viel spricht dafür, dass die Schulen im Rahmen eines institutionellen Lernprozesses die Bedingungen für das Verfassen der Maturaarbeit in den vergangenen Jahren weiter optimiert haben. Gleichwohl wird auch in unserer Studie deutlich, dass vielerorts durchaus noch ein beträchtlicher Handlungsspielraum für die weitere Verbesserung der Rahmenbedingungen besteht. Falls etwa ein höherer Anteil an interdisziplinären Maturaarbeiten erreicht werden soll, wäre es angeraten, diese Möglichkeit sowohl in den schriftlichen Vorgaben wie auch im Kontext der Betreuung stärker zu thematisieren.

5 Die wissenschaftspropädeutische Qualität der Maturaarbeiten

Um die wissenschaftspropädeutische Qualität der Maturaarbeiten beurteilen zu können, wurden die Arbeiten von je zwei Fachexperten gelesen und hinsichtlich der im Kriterienraster festgelegten Aspekte beurteilt. Allerdings liegt es in der Natur der Bewertungen, dass unterschiedliche Personen unterschiedliche Aspekte bei ihren Beurteilungen stärker oder weniger stark berücksichtigen (vgl. Ingenkamp, 1989). Durch ein gut strukturiertes Kriterienraster und eine intensive Schulung der Rater kann dieses Problem eingeschränkt, aber nie ganz behoben werden. Als Untersuchungsgegenstand stellen die Maturarbeiten noch aus weiteren Gründen eine besondere Herausforderung dar. So müssen bei ihrer Beurteilung nicht nur die verschiedensten Qualitätskriterien berücksichtigt, sondern es muss ebenfalls versucht werden, Traditionen und Konventionen unterschiedlicher Fachgebiete mit einem einzigen Instrument zu erfassen. Eine Vielzahl von wichtigen Aspekten entzieht sich dabei der exakten Quantifizierung und macht es den Ratern schwer, einheitliche Urteile zu fällen.

5.1 Erst- und Zweitratings

Wir sind davon ausgegangen, dass auch mit einer sehr intensiven Raterschulung Abweichungen zwischen den Beurteilungen der Rater nicht auszuschliessen sind. Diese Unterschiede können durch spezifische Eigenschaften der jeweiligen Maturaarbeit beeinflusst sein, die nicht alle im Detail in einem Kriterienraster Berücksichtigung finden. Es ist in solchen Fällen also nicht eindeutig zu klären, welche Gesichtspunkte mit in die Beurteilung einfliessen. Wir haben deshalb bei gravierenden Abweichungen zwischen den beiden Ratern die itemspezifischen Beurteilungen bei der Analyse der betreffenden Kriterien unberücksichtigt gelassen. Das geschah auch aufgrund der Erwägung, dass Berechnungen auf der Grundlage des arithmetischen Mittels bei einer Streuung von mehr als zwei Skalenpunkten einen irreführenden Anschein von Homogenität erzeugen. Der Prozentsatz der Beurteilungskriterien einzelner Maturaarbeiten bei denen dies jeweils gilt, ist relativ gering, wie aus Tabelle VI.24 entnommen werden kann.

Tab. VI.24: Anteil der Ratingdifferenzen ≤ 2 Skalenpunkte (nach Kriterium, in Prozent)

Kriterium	Raterdifferenz ≤ 2 (Prozent)	Kriterium	Raterdifferenz ≤ 2 (Prozent)
Globalurteil	94,1	Reflexion Arbeitsprozess	83,3
Titelblatt	98,2	Korrektes Zitieren	85,5
Layout	97,5	Quellenangaben	85,9
Tabellen Abbildungen	90,9	Angemessenheit Quellen	90,4
Textfluss	91,8	Umgang Tabellen Abbild	87,7
Prägnanz	93,4	Fragestellung allgemein	84,6
Wortwahl	93,6	Formulierung Fragest	84,6
Ausdruck	93,6	Einbettung Thema	86,5
Satzstruktur	92,4	Überblick Aufbau	81,2
Sprachliche Korrektheit	93,1	Differenzierung	86,8
formale Gliederung	94,3	Wertfreiheit	86,8
Titel	94,0	Schlussfolgerung	87,3
Feingl. Titelhierarchie	91,3	Angemessenheit Methoden	88,7
Umfang und Gewichtung	91,8	Beschreibung Methoden	87,2
Schwerpunktsetzung	89,9	Begrifflichkeiten	85,9
Kohärenz Argumentation	90,8	Verwendung der Begriffe	90,3

Wegen der Komplexität des Themas war es nötig, eine beträchtliche Anzahl von Fachexpertinnen und Fachexperten mit unterschiedlichem schulsprachlichem Hintergrund einzubinden. Die Messung der Interrater-Reliabilität wird dadurch beträchtlich erschwert, dass diese Fachpersonen in verschieden zusammengesetzten Dyaden tätig waren und teilweise nur wenige Fälle zu beurteilen hatten. Für die Intraklassenkorrelation ergibt sich beim Globalurteil, das den Gesamteindruck der Rater zu jeweils einer Maturaarbeit widerspiegelt, mit ICC=.732 ein solider Wert. In dieser Korrelation bestätigen sich die hohen Prozentwerte der Einschätzungen, in denen die Rater nicht mehr als zwei Skalenwerte voneinander abweichen, so dass wir insgesamt von einer zufriedenstellenden Übereinstimmung sprechen können.

5.2 Angaben zur Stichprobe

Die Analyse der wissenschaftspropädeutischen Qualität der Maturaarbeiten basiert auf einer Stichprobe von 437 Arbeiten; davon entfallen 56,8% auf die Deutschschweiz, 35,5% auf die Romandie und 7,8% auf den Kanton Tessin. 380 der ausgewerteten Arbeiten sind als Einzelarbeiten erstellt worden, 42 Arbeiten in einer Zweiergruppe und 15 Arbeiten in einem Team von drei Personen.

Die Grobeinteilung der Stichprobe nach Fachgebieten ergibt folgendes Bild: Rund zwei Drittel der Maturaarbeiten fallen in den sozialwissenschaftlichen Bereich, zu dem neben Geographie und Geschichte auch Themengebiete wie Wirtschaft, Recht und Politik gehören. Knapp 15% sind nach Einschätzung der Rater dem geisteswissenschaftlichen Bereich zuzuordnen, rund 17% dem naturwissenschaftlich-mathematischen Bereich. Bei der hier vorgenommenen Zuordnung werden die Sprachen mit zum geisteswissenschaftlichen Bereich gerechnet.

Tab. VI.25: Zuordnung nach Fachgebieten

Fachgebiet	Anteile in Prozent
Geisteswissenschaften	14,8
Sozialwissenschaften	68,1
Naturwissenschaften / Mathematik	17,1

55,6% der Arbeiten werden von den Ratern einhellig als monodisziplinär eingestuft, 17,6% ebenso einhellig als interdisziplinär.[59] In rund einem Viertel der Fälle besteht keine Einigkeit darüber, ob die Arbeit einem oder mehreren Fachgebieten zuzurechnen ist. Ein Drittel der Arbeiten wird als überwiegend theoretisch eingestuft, rund 30% gelten als gleichermassen theoretische und empirische Arbeiten, knapp 12% als überwiegend empirisch. Auch hier ist die Anzahl der Fälle beträchtlich, bei denen die Rater nicht zu einem gleich lautenden Urteil kommen: Bei einem Viertel der Fälle weicht die Einschätzung der Expertinnen und Experten im Hinblick auf das Verhältnis von Theorie und Empirie voneinander ab. Hier deutet sich an, dass das Kriterienraster hinsichtlich der Fragen der Einordnung zu undifferenziert bleibt und eventuell eine Einschätzung des jeweiligen Anteils an theoretischen vs. empirischen und monodisziplinären vs. interdisziplinären Projektkomponenten ergiebiger gewesen wäre.

In Bezug auf den Plagiatsverdacht zeigt sich, dass das Raterurteil eine relativ hohe Übereinstimmung mit der Plagiatssoftware *Plagiarism Finder* (siehe auch Kap. 2.2.5) aufweist: Von sieben des Plagiats verdächtigten Arbeiten, welche dem Test mit *Plagiarism Finder* unterzogen wurden, wiesen zwei keine Übereinstimmung mit Texten aus dem Internet auf, vier Arbeiten zeigten Übereinstimmungen zwischen 2% und 5%, eine Arbeit 13%. Bei sechs als unverdächtig bewerteten Arbeiten, welche als Kontrolle ebenfalls mittels der Plagiatssoftware geprüft wurden, bestätigt sich tatsächlich in keinem Fall, dass fremdes geistiges Eigentum ohne Kennzeichnung übernommen wurde: Die Software weist für diese Arbeiten durchweg 0% übereinstimmende Textstellen aus.

Im Gegensatz zur Überprüfung mittels Plagiatssoftware zeigte aber die Überprüfung mittels „Google-Technik“ ein anderes und vor allem differenzierteres Bild. Der zusätzliche stichprobenartige Vergleich einzelner Textteile

[59] Allerdings ist anzumerken, dass die Einstufung zum mono- oder interdisziplinären Charakter der Arbeiten nicht immer eindeutig vorgenommen werden konnte. Häufig handelt es sich um monodisziplinäre Arbeiten, die kleine Anteile aus anderen Fachgebieten aufgreifen. Eine disziplinär klare Grenze zu ziehen, hat sich im Laufe der Bewertungsarbeit teils als schwierig herausgestellt. Es zeigt sich weiterhin nur eine geringe Übereinstimmung zwischen der Einschätzung der Rater zur Mono- bzw. Interdisziplinarität mit den Angaben der Maturandinnen und Maturanden, in wie vielen Haupt- und Nebenfächern sie ihre Arbeit geschrieben haben.

der ausgewählten Maturaarbeiten mit Texten, die über die Suchmaschine Google gefunden wurden, machte unmittelbar deutlich, dass es sich nicht bei allen durch die Software gekennzeichneten auch um plagiierte Textstellen handelt. Von den 20 speziell untersuchten Arbeiten fanden sich in neun Arbeiten Textstellen, die wortwörtlich ohne entsprechende Kennzeichnung aus dem Internet übernommen wurden. Insgesamt sieben dieser Arbeiten wurden auch durch einen oder beide Rater des Plagiats verdächtigt, in zwei Fällen schöpften sie keinen Verdacht. Elf Arbeiten stellen sich nach dieser Suchmethode als plagiatsfrei heraus. In fünfzehn Arbeiten wurde zudem nichtalltägliches Wissen präsentiert, ohne dass im Text explizit auf die Quelle verwiesen worden wäre, in den meisten dieser Fälle wurden die Quellen aber immerhin im Literaturverzeichnis aufgeführt. Bis auf eine Ausnahme zeigt sich in all diesen Arbeiten, dass nicht durchgehend korrekt zitiert wurde. Sehr oft werden etwa aus anderen Werken übernommene Gedanken nicht direkt gekennzeichnet, sondern die Quellen ohne klaren inhaltlichen Bezug nur am Anfang oder Ende eines längeren Abschnittes oder Kapitels angegeben. Häufig werden auch Zitate nicht durch Anführungsstriche als solche gekennzeichnet. Ebenso werden Internetquellen in den meisten Arbeiten nur mit der URL belegt, ohne aber einen Autor oder eine Autorin sowie den Titel der entsprechenden Website zu nennen.

Hinsichtlich des Plagiatsverdachts ist ausdrücklich zu betonen, dass hier letztlich die Täuschungsabsicht entscheidend ist: Ein Plagiat liegt dann vor, wenn jemand sich „mit fremden Federn schmückt“, indem er die Erkenntnisse und Errungenschaften anderer unter bewusster Vertuschung oder Verleugnung der Urheberschaft als eigene Leistungen ausgibt. Bei vielen der hier beanstandeten Arbeiten ist indes zu vermuten, dass das Problem tatsächlich durch handwerkliche Mängel und nicht durch betrügerische Absichten verursacht ist.

5.3 Skalenbildung

Die Beurteilung der Qualität der Maturaarbeiten basiert auf insgesamt 32 Items, die sich auf formale, sprachliche und inhaltliche Aspekte sowie ein Globalurteil zum Gesamteindruck der Rater erstrecken. Durch eine Faktorenanalyse lassen sich hier drei Skalen mit soliden Reliabilitätswerten erhärten.

Tab. VI.26: Faktorladungen zu den Qualitätsmerkmalen der Maturaarbeiten

Kriterium	Faktor 1	Faktor 2	Faktor 3
Titelblatt			0,49
Layout			0,51
Optik von Tabellen & Abbildungen			0,82
Textfluss		0,81	
Prägnanz	0,45	0,76	
Wortwahl	0,43	0,76	
Ausdruck		0,83	
Satzstruktur		0,85	
Sprachliche Korrektheit		0,79	
formale Gliederung	0,55		0,63
Titel	0,58		
Feingliederung Titelhierarchie	0,70		
Umfang und Gewichtung	0,77		

Kriterium	Faktor 1	Faktor 2	Faktor 3
Schwerpunktsetzung	0,77	0,40	
Kohärenz der Argumentation	0,83		
Reflexion des Arbeitsprozesses	0,77	0,40	
Korrektes Zitieren			0,67
Quellenangaben			0,61
Angemessenheit der Quellen	0,55		0,54
Funktionalität von Tabellen & Abb.			0,83
Fragestellung allgemein	0,75		
Formulierung Fragestellung	0,69		0,40
Einbettung Thema	0,48	0,40	
Überblick Aufbau	0,73		
Differenzierung	0,63		
Wertfreiheit	0,57		0,42
Schlussfolgerung	0,67	0,49	
Angemessenheit der Methoden	0,84		
Beschreibung der Methoden	0,69		
Begrifflichkeiten	0,67		
Verwendung der Begriffe	0,57	0,59	

(Faktorladungen >.4, varimax-rotiert)

Auf dem ersten Faktor, der rund 33% der Gesamtvarianz erklärt, laden die Items *Angemessenheit der Methoden* und *Kohärenz der Argumentation* am höchsten. Beim zweiten Faktor, der im Umfang von knapp 20% zur Erklärung der Gesamtvarianz beiträgt, stechen *Ausdruck* und *Satzstruktur* sowie *Textfluss* und *sprachliche Korrektheit* hervor. Beim dritten Faktor, der rund 15% der Gesamtvarianz erklärt, laden die Items zur Optik und zur Funktionalität von Tabellen und Abbildungen am höchsten. Mehrere Items laden relativ stark auf zwei Faktoren: Das gilt für die Prägnanz des sprachlichen Ausdrucks und die Wortwahl, die thematische Schwerpunktsetzung, die Reflexion des Arbeitsprozesses, die Einbettung des Themas, die Schlussfolgerungen und die Verwendung der Terminologie.

Faktor I: Die Skala *Inhaltliche Qualität* umfasst 11 Items, die sich auf theoretische und methodologische Elemente wie auch auf die thematische Fokussierung und die argumentative Stringenz der Maturaarbeit erstrecken. Bei der Bildung dieser Skala sind mehrere Items ausgeschlossen worden, die sowohl auf diesem wie auch auf anderen Faktoren relativ hoch laden. Auf den ersten Blick mag es irritierend erscheinen, dass die Methodik hier unter den Inhalt subsumiert wird, tatsächlich trägt aber die Wahl und der Einsatz angemessener Methoden ganz entscheidend zum Sachgehalt wissenschaftspropädeutischer und wissenschaftlicher Arbeiten bei.

Tab. VI.27: Skala "Inhaltliche Qualität"

Item	Kennwerte	Antwortmöglichkeiten
Titel Feingliederung Titelhierarchie Umfang und Gewichtung Schwerpunktsetzung Kohärenz der Argumentation Fragestellung allgemein Überblick Aufbau Differenzierung Überblick Aufbau Beschreibung der Methoden Begrifflichkeiten	gültige Fälle: n=437 Cronbach's Alpha: .93 M=2,7 SD=0,92	1 Anforderungen voll erfüllt 2 3 4 5 6 Anforderungen überhaupt nicht erfüllt

Faktor II: Die Skala *Sprachliche Qualität* umfasst sechs Items, die sich auf Ausdruck und Stil wie auch auf Grammatik und Rechtschreibung erstrecken und mithin den Grad der korrekten und kohärenten Verschriftlichung von Ergebnissen abbilden.

Tab. VI.28: Skala "Sprachliche Qualität"

Item	Kennwerte	Antwortmöglichkeiten
Textfluss Prägnanz Wortwahl Ausdruck Satzstruktur Sprachliche Korrektheit	gültige Fälle: n=432 Cronbach's Alpha: .94 M=2,5 SD=0,91	1 Anforderungen voll erfüllt 2 3 4 5 6 Anforderungen überhaupt nicht erfüllt

Das Item *Verwendung der Begriffe*, das sowohl auf diesem Faktor als auch auf dem Faktor I relativ hoch lädt, passt wegen seiner semantischen Fokussierung sachlogisch nicht in diese Skala hinein und trägt auch tatsächlich mit einem Trennschärfekoeffizient von 0,57 nicht zur Skalenqualität bei.

Tab. VI.29: Skala "Formale Qualität"

Item	Kennwerte	Antwortmöglichkeiten
Titelblatt Layout Optik von Tabellen & Abb. Formale Gliederung Korrektes Zitieren Quellenangaben Funktionalität von Tab. & Abb.	gültige Fälle: n=437 Cronbach's Alpha: .75 M=2,5 SD=0,79	1 Anforderungen voll erfüllt 2 3 4 5 6 Anforderungen überhaupt nicht erfüllt

Faktor III: Die Skala *Formale Qualität* umfasst sieben Items, die das Erscheinungsbild der Maturaarbeit unter dem Aspekt der Textgestaltung erfassen. Dies beschränkt sich nicht allein auf den visuellen Eindruck, den das schriftliche Produkt beim Rezipienten hinterlässt: Neben der Lesbarkeit und Übersichtlichkeit geht es nicht minder um die Einhaltung der Regeln des wissenschaftlichen Diskurses durch die korrekte Bezugnahme auf fremdes Gedankengut, das bei der Durchführung des Projekts verwendet wurde.

Durch die Faktorenanalyse lassen sich Hintergrundvariablen ermitteln, die sich auf verschiedene Merkmalsdimensionen erstrecken. Diese Hintergrundvariablen bilden mithin heterogene Qualitätsaspekte ab, wobei die Ausprägungen dieser Merkmale aber durchaus miteinander korrelieren können. In der Tat zeigt sich, dass alle drei Skalen hochsignifikant positiv miteinander korrelieren: Bei einer Arbeit mit hoher inhaltlicher Qualität ist die Wahrscheinlichkeit beträchtlich, dass diese Arbeit auch sprachlich und formal überdurchschnittlich sein wird.

Tab. VI.30: Korrelationen der drei Skalen "Inhaltliche Qualität", "Sprachliche Qualität" und "Formale Qualität"

	Inhaltliche Qualität	Sprachliche Qualität	Formale Qualität
Inhaltliche Qualität		0,67**	0,6**
Sprachliche Qualität	0,67*		0,53**
Formale Qualität	0,6**	0,53**	

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

5.4 Inhaltliche Qualität

Die Skala *Inhaltliche Qualität* bezieht sich auf die thematische Fokussierung, die methodologische Fundierung und die argumentative Schlüssigkeit der Maturaarbeiten, wobei nicht zwischen theoretisch oder empirisch ausgerichteten Projekten unterschieden wird: Die Forderung nach klar formulierten Fragestellungen, nach sinnvoll ausgewählten und eingesetzten Methoden sowie nach einer folgerichtigen Gedankenführung gilt unabhängig von allen Disziplinen und auch unabhängig von jeder fachlichen Binnendifferenzierung.

Die Skala weist ein arithmetisches Mittel von M=2,7 bei einer Standardabweichung von SD=0,92 auf. Mit Blick auf die einzelnen Items dieser Skala ist festzustellen, dass die Titelgebung in der Regel triftig ist und dass die Feingliederung der gesamten Arbeit recht gut gelingt. Auch wissen die meisten Schülerinnen und Schüler in ihren Arbeiten eindeutige Schwerpunkte zu setzen und den Text durchgängig auf das gesetzte Thema zu beziehen. Entsprechend werden auch die verwendeten Fachbegriffe definiert und kontextgerecht eingesetzt. Etwas weniger positiv fallen die Bewertung der Rater bezüglich des differenzierten Argumentierens aus. Nicht immer gelingt es den Schreibenden, verschiedene Ansätze zur Beantwortung der Fragestellung einzubringen oder die von ihnen gewählte Perspektive auf dem Hintergrund einer Abwägung zu begründen.

Überraschendes zeigt sich bei der Wahl der Methoden und deren Beschreibung: Während die Rater die verwendeten Methoden in der Regel als dem Ziel und dem Thema angemessen beurteilt haben, fällt das Urteil hinsichtlich der Beschreibung dieser Methoden kritischer aus: Die Schülerinnen und Schüler tun häufig das Richtige, aber sie rekonstruieren das eigene Handeln im Projektbericht nicht immer so, dass das Vorgehen der intersubjektiven Überprüfung zugänglich wäre.

Tab. VI.31: Mittelwerte und Standardabweichungen für die Kriterien der Skala "Inhaltliche Qualität"

Kriterium		1 2 3 4 5 6
Titel	2.1 (0.97)	
Feingliederung Titelhierarchie	2.5 (0.99)	
Umfang und Gewichtung	2.8 (1.06)	
Schwerpunktsetzung	2.5 (1.08)	
Kohärenz der Argumentation	2.8 (1.13)	
Fragestellung allgemein	2.6 (1.33)	
Überblick Aufbau	2.9 (1.61)	
Differenzierung	3.1 (1.35)	
Angemessenheit Methoden	2.4 (1.18)	
Beschreibung der Methoden	3.4 (1.56)	
Begrifflichkeiten	2.3 (1.10)	

Eine Varianzanalyse zeigt, dass sich im Mehrfachvergleich für die Geistes-, Sozial- und Naturwissenschaften keine signifikanten Differenzen bei der inhaltlichen Qualität der Arbeiten feststellen lassen. Während der t-Test beim Geschlechtervergleich ebenfalls keine signifikanten Werte ergibt, schneiden die Gruppenarbeiten mit einem Mittelwert von 2,4 signifikant ($\alpha<0{,}001$, zweiseitig) besser ab als die Einzelarbeiten (M=2,7). Hier gilt vermutlich, dass bei den Gruppenarbeiten die inhaltlichen Aspekte, von mehreren Personen diskutiert und durchdacht werden, bevor sie in die Aufnahme in die Maturaarbeit finden. Die Varianzanalyse zur regionalen Ausprägung der inhaltlichen Qualität ergibt, dass die Schülerinnen und Schüler aus der Romandie (M=2,1) signifikant ($\alpha<0{,}001$, zweiseitig) besser abschneiden als ihre Peers in der Deutschschweiz (M=3,0); für den Kanton Tessin ergibt sich ein Durchschnittswert von M=2,7, wobei es sich aber aufgrund der Fallzahl empfiehlt, auf einen Signifikanztest zu verzichten. Im Bezug auf die sprachregionalen Unterschiede erweist es sich als misslich, dass es nicht gelungen ist, im grösseren Umfang bilinguale Fachpersonen zu rekrutieren, deren Beteiligung am Rating in zwei Landesteilen einen direkten varianzanalytischen Vergleich zwischen den Regionen ermöglichen würde: Trotz der Einheitlichkeit des Instrumentariums und trotz des einheitlichen Schulungsverfahrens ist nicht auszuschliessen, dass hier unterschiedliche Lehr- und Lernkulturen mit unterschiedlichen Bewertungsmassstäben einhergehen.

5.5 Sprachliche Qualität

Die Skala *Sprachliche Qualität* umfasst zum einen Items, die sich eher auf Fragen des Stils beziehen, und zum anderen auf Items, bei denen es um die korrekte Schreibweise und den regelkonformen Satzbau geht. Hier ist mithin dem Umstand Rechnung zu tragen, dass zwei verschiedene Aspekte unterschiedlicher Gewichtigkeit im Spiel sind: Während der fehlerfreie Umgang mit Grammatik und Rechtschreibung in allen Wissenschaftsgebieten unabdingbar ist, kommt stilistischen Aspekten in geisteswissenschaftlichen Texten eine weit höhere Bedeutung zu als in anderen Wissenschaftsgebieten: Unter naturwissenschaftlich-technischen und mathematischen Gesichtspunkten kann es zwingend geboten sein, zugunsten der Präzision auf stilistische Feinheiten zu verzich-

ten – insbesondere auf eine abwechslungsreiche Diktion. Unter diesem Gesichtswinkel ist es auch hier wichtig, der disziplinären Ausrichtung der Maturaarbeiten ein besonderes Augenmerk zu schenken.

Die Skala Sprachliche Qualität weist ein arithmetisches Mittel von M=2,5 bei einer Standardabweichung von SD=0,91 auf. Mit Blick auf die Einzelitems dieser Skala ist festzuhalten, dass die Bewertung der Rater durchschnittlich im positiven Bereich liegen. Weder bei der sprachlichen Verständlichkeit noch beim Textfluss gibt es massive Beanstandungen. Ebenso gut steht es um Orthographie, Grammatik und Interpunktion. Bezogen auf die Mittelwerte schneidet der sprachliche Ausdruck am wenigsten positiv ab: Beschreibungen, Analysen und Reflexionen sind des Öfteren von umgangssprachlichen alltagsnahen Wendungen durchsetzt.

Tab. VI.32: Mittelwerte und Standardabweichungen für die Kriterien der Skala "Sprachliche Qualität"

Kriterium		1	2	3	4	5	6
Textfluss	2.2 (0.91)						
Prägnanz	2.6 (1.05)						
Wortwahl	2.5 (1.01)						
Ausdruck	2.7 (1.10)						
Satzstruktur	2.6 (0.96)						
Sprachliche Korrektheit	2.3 (1.01)						

Die Varianzanalyse ergibt hier im Mehrfachvergleich, dass die sprachliche Qualität naturwissenschaftlicher Arbeiten (M=2,2) signifikant ($\alpha<0{,}05$) höher beurteilt wird als die der sozialwissenschaftlichen Arbeiten (M=2,6). Für die geisteswissenschaftlichen Arbeiten ergeben sich hier in beiderlei Hinsicht keine signifikanten Abweichungen. Es ist nicht auszuschliessen, dass die hoch bewertete sprachliche Qualität der naturwissenschaftlichen und mathematischen Arbeiten auf die unterschiedlichen Ansprüche, die von den jeweiligen Fachexperten an eine wissenschaftspropädeutische Arbeit gestellt werden, zurückzuführen ist.

Ein t-Test im Hinblick auf den Genderaspekt ergibt keine signifikanten Unterschiede. Beim Vergleich von Einzel- und Gruppenarbeiten ergibt sich ebenfalls keine signifikante Abweichung. Die Varianzanalyse zur regionalen Ausprägung der sprachlichen Qualität ergibt auch hier, dass die Schülerinnen und Schüler aus der Romandie (M=2,1) signifikant ($\alpha<0{,}001$) besser abschneiden als ihre Peers in der Deutschschweiz (M=2,8). Für den Kanton Tessin ergibt sich ein ähnlicher Durchschnitt wie für die Romandie (M=2,2).

5.6 Formale Qualität

Für die schriftliche Präsentation von wissenschaftlichen Ergebnissen haben sich Konventionen etabliert, die von Fachgebiet zu Fachgebiet variieren. Augenfällig wird das etwa, wenn man Literaturverzeichnisse in natur- oder ingenieurwissenschaftlichen Fachzeitschriften mit denen von geisteswissenschaftlichen Fachzeitschriften vergleicht. Bei näherem Hinsehen stellt sich dann in der Regel schnell heraus, dass auch die Zitate und die Quellenverweise im Text verschiedenen Konventionen gehorchen. Gleichwohl wird aber innerhalb der Fachgebiete auf die Einheitlichkeit der Präsentationsform insistiert; und dies gilt auch dann, wenn – etwa im Bereich der Sozialwissenschaften – die Vorgaben für die Zitierweise und die Literaturnachweise von Journal zu Journal voneinander abweichen. Im Rahmen des Ratings der Qualität der Maturaarbeiten geht es nicht darum, die Vorzüge und Nachteile der domänenspezifischen Konventionen gegeneinander abzuwägen, sondern es geht allein darum, die Einheitlichkeit der Handhabung formaler Vorgaben zu beurteilen.

Die Skala *Formale Qualität* weist ein arithmetisches Mittel von M=2,5 bei einer Standardabweichung von SD=0,79 auf. Der Mittelwert liegt damit im Bereich positiver Bewertungen und zeigt an, dass die wissenschaftspropädeutischen Anforderungen an die Textgestaltung überwiegend erfüllt werden. Die einzelnen Werte der Skala zur formalen Qualität der Arbeiten zeigen, dass das Titelblatt und die formale Textgliederung von den Ratern in der Regel als gut bis sehr gut bewertet wurden: Mehrheitlich können den Arbeiten auf den ersten Blick die wichtigsten Informationen zum Sujet, zu den Autoren oder Autorinnen sowie zum Kontext und Zeitpunkt der Erstellung abgelesen werden. Diese positiven Werte sind nicht zuletzt darauf zurückzuführen, dass hier die meisten Schulen klare Gestaltungsvorgaben machen. Ähnlich verhält es sich hinsichtlich der formalen Unterteilung des Textes in Einleitung, Hauptteil und Schluss.

Knapp positiv fallen die Urteile zum Zitieren, zu den Quellenangaben und zum Umgang mit Tabellen sowie Abbildungen aus. Wie bereits oben erwähnt, bereitet den Schülerinnen und Schülern vor allem die korrekte Angabe von Quellen aus dem Internet erhebliche Mühe. Einerseits werden Textpassagen des Öfteren nur global gekennzeichnet, ohne genauer auf die zitierten Stellen hinzuweisen. Andererseits beschränken sich die Nachweise häufig auf die wortwörtliche Wiedergabe von Textpassagen in der Form des Zitats; bei Entlehnung aus Quellen oder aus der Sekundärliteratur in der Form der Paraphrase unterbleibt nicht selten die präzise Kennzeichnung. Darüber hinaus ist bemerkenswert, dass das Rating zur grafischen Textgestaltung nicht so positiv ausgefallen ist, wie es die Selbsteinschätzung der Schülerinnen und Schüler erwarten liesse: Im Umgang mit dem Computer, beispielsweise beim Festlegen des Schriftbildes, der Silbentrennung und dem Einsetzen der Seitenzahlen, steht es um die Anwendungskenntnisse der Maturandinnen und Maturanden mitunter weniger gut, als von ihnen selbst angenommen wird (vgl. Kap. 4.5).

Tab. VI.33: Mittelwerte und Standardabweichungen für die Kriterien der Skala "Formale Qualität"

Kriterium		1	2	3	4	5	6
Titelblatt	1.8 (0.98)						
Layout	2.5 (0.94)						
Optik von Tabellen & Abb.	2.8 (1.12)						
Formale Gliederung	2.0 (0.95)						
Korrektes Zitieren	2.9 (1.41)						
Quellenangaben	3.0 (1.56)						
Funktionalität von Tab. & Abb.	2.8 (1.21)						

Bei der Zuordnung der Maturaarbeiten zu Fachgebieten ergibt die Varianzanalyse im Mehrfachvergleich für die Geistes-, Sozial- und Naturwissenschaften keine signifikanten Unterschiede. Auf der Grundlage eines t-Tests ist festzustellen, dass die formale Qualität der Arbeiten nicht wesentlich dadurch beeinflusst wird, ob es sich um Einzelarbeiten oder um Gruppenarbeiten handelt. Auch unter dem Genderaspekt lassen sich hier keine signifikanten Unterschiede feststellen. Hingegen zeigt sich im sprachregionalen Vergleich, dass die Schülerinnen und Schüler aus der Romandie (M=2,3) bei der Beurteilung der formalen Qualität der Arbeiten abermals signifikant ($\alpha<0{,}001$, zweiseitig) besser abschneiden als ihre Peers in der Deutschschweiz (M=2,7).[60] Der Durchschnitt für die italienischsprachige Schweiz liegt bei der formalen Qualität bei M=2,8.

[60] Wegen der geringen Fallzahl bei einer vergleichsweise sehr grossen Menge von Fachpersonen, die hier das Rating durchgeführt hat, ist die Teilstichprobe für den Kanton Tessin nicht in diesen Regionalvergleich einbezogen worden.

5.7 Globalurteil

Das Globalurteil bildet den Gesamteindruck der Expertinnen und Experten zu jeweils einer Maturaarbeit ab. Bei der Raterschulung ist ausdrücklich betont worden, dass es sich nicht um eine Schulnote handelt, allerdings ist bei einer Skala von 1 bis 6 dieser Suggestion kaum zu entkommen, selbst wenn die Polung sich invers zur gängigen Praxis der Notenvergabe in der Schweiz verhält und die 6 mithin die schlechteste, die 1 dagegen die beste Bewertung darstellt.

Für das Globalurteil ergibt sich ein arithmetischer Mittelwert von 3,1 bei einer Standardabweichung von 1,16. Der Durchschnitt liegt damit knapp unterhalb des virtuellen Skalenmittelpunkts im schwach positiven Wertungsbereich; der Gesamteindruck ist von daher als zufrieden stellend zu bezeichnen.

Das Globalurteil korreliert mit den drei Skalen zur inhaltlichen, sprachlichen und formalen Qualität jeweils hoch. Eine lineare Regressionsanalyse zeigt, dass 79% der Varianz des Globalurteils auf die Einschätzungen zur inhaltlichen, sprachlichen und formalen Qualität zurückzuführen sind. Die inhaltlichen Aspekte beeinflussen das Globalurteil am stärksten, die formalen Aspekte hingegen am wenigsten. Der Beta-Koeffizient liegt für die inhaltliche Qualität bei 0,58, für die sprachliche Qualität bei 0,23 und für die formale Qualität bei 0,1.

Tab. VI.34: Korrelation Globalurteil mit den drei Skalen "Inhaltliche Qualität", "Sprachliche Qualität" und "Formale Qualität"

	Inhaltliche Qualität	Sprachliche Qualität	Formale Qualität
Globalurteil	0,85**	0,76**	0,62**

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Im Mehrfachvergleich von Geistes, Sozial- und Naturwissenschaften ergeben sich hinsichtlich des Globalurteils keine signifikanten Unterschiede; ein gleichermassen unauffälliger Befund ergibt sich bei der Durchführung eines t-Tests für den Genderaspekt. Gruppenarbeiten (M=2,7) schneiden beim Globalurteil signifikant ($\alpha<0{,}05$, zweiseitig) besser ab als Einzelarbeiten (M=3,1). Im Regionalvergleich ergibt sich auch für das Globalurteil, dass die Maturandinnen und Maturanden aus der Romandie (M=2,6) signifikant ($\alpha<0{,}001$, zweiseitig) besser abschneiden als jene aus der Deutschschweiz (M=3,4); für den Kanton Tessin ergibt sich hier ein Wert von M=2,9. Auch hier ist wieder der Vorbehalt zu machen, dass sich aufgrund der sprachregionalen Segregation der Rater nicht aufhellen lässt, ob und in welchem Umfang hier unterschiedliche Bewertungskulturen von Bedeutung sind.

5.8 Individuelle und institutionelle Qualitätsaspekte

Im Hinblick auf die Textverarbeitung und die grafische Gestaltung der Projektberichte haben wir bereits kurz angedeutet, dass die Einschätzung der Rater nicht unbedingt mit der Selbsteinschätzung der Maturandinnen und Maturanden kongruent ist. Im Folgenden soll systematisch beleuchtet werden, wie sich die Evaluation der Qualität der Maturaarbeiten zur Perspektive der Verfasserinnen und Verfasser verhält.

Themenwahl

Unserer Hypothese zufolge sollte die inhaltliche Qualität der Maturaarbeit mit dem Grad der Freiheit bei der Themenwahl steigen. Eine Varianzanalyse mit Bonferronikorrektur für Mehrfachvergleiche zeigt jedoch, dass im Gegenteil die inhaltliche Qualität der Arbeiten bei freier Themenwahl (M=2,8) signifikant ($\alpha<<0{,}05$) geringer ist als bei der Vorgabe der Themen durch eine Liste (M=2,2) oder durch ein Rahmenthema (M=2,4). Des Weiteren zeigt sich, dass sich auch die spezifische Unterstützung der Lehrperson bei der Wahl des Themas positiv auf die inhaltliche Qualität der Arbeit auswirkt. Eine Erklärung dafür könnte in der Betreuungssituation zu finden sein: Es erscheint plausibel, dass die Lehrperson auf dem Hintergrund einer vorstrukturierten und limitierten Themenfindung besser auf spezifische Fragen und Probleme zu reagieren vermag, auch wenn sich dies in der

Wahrnehmung der Befragten nicht so widerspiegelt. Maturandinnen und Maturanden mit gebundenen Themen nehmen die Betreuungssituation nicht anders wahr als diejenigen, die ihr Thema frei wählen konnten. Indes spricht die höhere inhaltliche Qualität der Arbeiten mit gebundenen Themen dafür, dass die Betreuung hier eine grössere Wirksamkeit entfalten konnte. Diese Interpretation wird auch dadurch unterstützt, dass es keinen signifikanten Zusammenhang zwischen der perzipierten Betreuungssituation und der durch das Rating diagnostizierten Qualität der Arbeiten gibt.

Die Limitierung der Themen kann sich zudem positiv auf die Zusammenarbeit zwischen den Schülerinnen und Schülern auswirken, weil die grössere Nähe der jeweils zu bearbeitenden Fragestellungen einen Austausch befruchtender macht als dies bei Themen der Fall wäre, die innerhalb der Klasse oder des Jahrgangs singulär sind.

Der relativ überraschende Befund, dass die Freiheit bei der Themenwahl keineswegs mit höherer Motivation und besseren Resultaten gleichzusetzen ist, unterstreicht die Wichtigkeit klar und verbindlich ausformulierter institutioneller Rahmenbedingungen. Die vorliegenden Befunde sprechen dafür, dass beides für die inhaltliche Qualität wichtiger ist als das Interesse oder das Vorwissen der Maturandinnen und Maturanden, welche keine signifikanten Zusammenhänge zur Qualität der Arbeiten aufweisen.

Zwischen natur-, sozial- und geisteswissenschaftlich ausgerichteten Maturaarbeiten gibt es gesamthaft kaum signifikante Qualitätsunterschiede. Zudem ist festzustellen, dass interdisziplinär ausgerichtete Arbeiten keine bedeutsamen Differenzen zu Arbeiten mit klarer disziplinärer Zuordnung aufweisen.

Prozedurale Aspekte

Ähnlich wie die individuell wahrgenommene Betreuungssituation keinen signifikanten Einfluss auf die Qualität der Arbeiten hat, ist auch für die Häufigkeit der Kontakte mit Betreuerinnen oder Betreuern kein signifikanter Zusammenhang mit den verschiedenen Qualitätsaspekten nachzuweisen. Allerdings gibt es einzelne Aspekte, die für die Schülerinnen und Schüler bei der Bearbeitung ihrer Arbeit besonders wichtig waren. Neben der Unterstützung bei der Wahl des Themas sind für die inhaltliche Qualität der Arbeit die Hilfestellungen bei der Nutzung von elektronischen Medien, Hilfsangebote zu Arbeitsmethoden, zur Einhaltung formaler Regeln und zum Verfassen des Berichts bedeutsam. Die Hilfestellungen zum Verfassen der Arbeit wirken sich nicht nur auf die inhaltliche, sondern auch auf die sprachliche Qualität aus, zudem macht sich eine entsprechende Unterstützung auch positiv bei der formalen Qualität der Arbeiten bemerkbar. Generell sind die Zusammenhänge aber sehr schwach.

Tabelle VI.35 stellt die Korrelationen zu den prozeduralen Aspekten dar. Die Unterstützung wurde mit unterschiedlichen Items erhoben, deren Antwortalternativen sich auf einer vierstufigen Skala von 1 (sehr hilfreich) bis 4 (gar nicht hilfreich) abstufen. Diese Skala wurde hier in Entsprechung zum Raterbogen umgepolt.

Tab. VI.35: Korrelation der perzipierten Unterstützung durch die betreuenden Lehrperson mit den Qualitätsskalen

	Inhaltliche Qualität	Sprachliche Qualität	Formale Qualität
Wahl des Themas	0,14*	n.s.	n.s.
Fragestellung	n.s.	n.s.	n.s.
Zeitplanung	n.s.	n.s.	n.s.
Computernutzung	0,14*	n.s.	n.s.
Inhalt (fachlich)	n.s.	n.s.	n.s.
Arbeitsmethoden	0,11*	n.s.	n.s.
Aufbau / Gliederung	n.s.	n.s.	n.s.
Verfassen der Arbeit	0,14*	0,16**	n.s.
Motivation	n.s.	n.s.	n.s.
Formale Grundlagen	0,15**	0,13*	0,15**

* Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Wenn Schülerinnen und Schüler Vorbereitungsveranstaltungen besucht haben, fällt die Qualität der Maturaarbeiten in der Regel etwas besser aus. Dies gilt für die inhaltliche und die sprachliche Qualität, nicht aber für die formale Qualität der Arbeiten. Insbesondere im Zusammenhang mit Orientierungsveranstaltungen und Lernen am Projekt zeigen sich signifikante Zusammenhänge. Lernen am Projekt ist jedoch eine spezielle Veranstaltung, die nur die wenigsten Schülerinnen und Schüler in Anspruch nehmen konnten, in unserer Stichprobe waren es 32 Personen. Es zeigt sich an diesen Ergebnissen abermals, dass eine formell organisierte Vorbereitung und Begleitung der Maturaarbeiten sich als nützlich erweist. Schülerinnen und Schüler, die angeben, keine Vorbereitungsveranstaltung besucht zu haben, diese aber für wichtig erachtet hätten, erreichen in ihren Arbeiten eher eine geringere sprachliche Qualität, jedoch keine signifikant schlechteren Resultate bei der formalen und inhaltlichen Qualität. Die Vergleichsgruppe sind dabei jene Befragten, die keine Vorbereitungsveranstaltung besucht haben und das auch nicht als Nachteil ansehen.

Die Zeit, die die Schülerinnen und Schüler für ihre Arbeit aufbringen, steht mit der Qualität der Maturaarbeiten in keinem Zusammenhang. Dies gilt sowohl für die Bearbeitungszeit, die die Schülerinnen und Schüler im Fragebogen angegeben haben, als auch für die Zeit, die laut den Rahmenrichtlinien der Schulen vorgegeben sind. Ein Zusammenhang zwischen der Qualität der Arbeiten und den angegebenen Vorkenntnissen der Schülerinnen und Schüler zu dem jeweiligen Thema besteht ebenfalls nicht. Allerdings haben Maturandinnen und Maturanden, die den Aufwand für ihre Arbeit im Vergleich zu anderen Anforderungen als zu hoch einstufen, mit ihren Projektberichten bei der Evaluation eher schwächer abgeschnitten (r=0,13** mit der sprachlichen Qualität und r=0,11* mit der inhaltlichen Qualität).

Die Verfügbarkeit eines Leitfadens und die wahrgenommene Transparenz der Abläufe wirken sich nicht positiv auf die Qualität der Maturaarbeit aus, diese scheint weit mehr von der unmittelbaren pädagogischen Betreuung beeinflusst zu sein. Auch hinsichtlich der wahrgenommenen Transparenz der Bewertungskriterien ist kein positiver Einfluss auszumachen. Vielmehr wurde die sprachliche Qualität der Arbeiten, deren Verfasser angaben, die Beurteilungskriterien vorab nicht gekannt zu haben, tendenziell eher höher eingestuft (r=0,15**). Zudem besteht ein schwacher negativer Zusammenhang zwischen der Einschätzung der Schülerinnen und Schüler zur Sinnhaftigkeit der Kriterien und der sprachlichen sowie der inhaltlichen Qualität der Arbeiten (r=-0,17** mit der sprachlichen Qualität und r=-0,11* mit der inhaltlichen Qualität). Werden die Kriterien explizit als nicht sinnvoll

angesehen, ist tendenziell auch die Qualität der Arbeit niedriger. Dieser Befund kann jedoch im Nachhinein dadurch zustande gekommen sein, dass Maturandinnen und Maturanden, die mit ihrer Maturaarbeit weniger erfolgreich waren, auch der Güte der Bewertungskriterien kritischer gegenüber stehen.

Neben der Unterstützung durch die Lehrperson spielt auch die Unterstützung durch andere Personen eine Rolle bei der Qualität der Maturaarbeit. Personen, die eine Unterstützung von ihren Geschwistern erhalten haben, erreichen in der Regel mit ihrer Maturaarbeit eine höhere formale Qualität (r=0,11*) und Personen, die eine Unterstützung durch Kollegen an der Schule hatten, erreichen eine höhere inhaltliche Qualität (r=0,1*). Auch hier handelt es sich unverkennbar um schwache, aber doch signifikante Korrelationen. Hingegen hat die von den Schülerinnen und Schülern deklarierte Unterstützung durch die Eltern keinen signifikanten Auswirkungen auf die Qualität der Arbeiten. Hier ist allerdings anzumerken, dass die Aussagekraft dieser Selbstdeklaration nicht sehr gross ist: Vollkommen ausser Betracht bleibt der latente Einfluss der Eltern auf die Bildungsbiographie.

Wirkung der Maturaarbeit

Nach Angaben der Maturandinnen und Maturanden hat sich die Beschäftigung mit der Maturaarbeit in der Regel positiv auf verschiedene Aspekte der individuellen Arbeitsorganisation und der kognitiven Kompetenz ausgewirkt (vgl. Kap. 4). Diejenigen Schülerinnen und Schüler, deren Arbeiten eher eine hohe sprachliche Qualität aufweisen, bezeugen einen positiven Einfluss des Projekts auf ihre kognitiven Kompetenzen (r=0,12*). Neben diesem sehr schwachen Zusammenhang zeigen sich keine signifikanten Effekte zwischen der Qualität der Arbeiten und den wahrgenommenen Wirkungen auf die individuelle Arbeitsorganisation und die kognitiven Kompetenzen. Die Benotung der Projektergebnisse steht hingegen in einem engerem Zusammenhang mit der perzipierten Wirkung der Maturaarbeit (r=0,23** mit der Wirkung auf die kognitive Kompetenz und r=0,3** mit der Wirkung auf die selbstständige Arbeitsorganisation). Es zeigt sich erneut, dass die Einschätzungen der Schülerinnen und Schüler relativ stark durch die institutionelle Bewertung ihrer Maturaarbeit beeinflusst sind.

Ein leichter Zusammenhang besteht zwischen der sprachlichen und inhaltlichen Qualität der Arbeiten und dem von den Verfasserinnen und Verfassern deklarierten Interessenzuwachs (jeweils r=0,11*). Sehr viel stärker korreliert dieser Zuwachs mit der Note (r=0,3**). Auch hinsichtlich der Frage, ob die Arbeit Spass gemacht hat, zeigt sich ein relativ starker Zusammenhang mit der Note (r=0,34**), aber kein signifikanter Zusammenhang mit dem Befund der Evaluation. Ob hier nachträglich die Wahrnehmung des eigenen Interesses und der Freude an der Arbeit durch den Erfolg beeinflusst wurde, oder andersherum das hohe Interesse und die Motivation der Schülerinnen und Schüler Eingang in die Note gefunden hat, lässt sich anhand der uns vorliegenden Daten nicht klären.

Auffällig ist, dass Schülerinnen und Schüler, deren Arbeiten im Rating eine überdurchschnittliche Qualität attestiert wird, den Nutzen für das Studium, die berufliche Ausbildung und die persönliche Entwicklung höher einschätzen. Diese Einschätzung erfolgte jeweils auf einer vierstufigen Skala von 1 = "sehr gering" bis 4= "sehr hoch"; die Korrelationen sind in Tabelle VI.36 dargestellt.

Tab. VI.36: Korrelation des eingeschätzten Nutzens mit den drei Skalen "Inhaltliche Qualität", "Sprachliche Qualität" und "Formale Qualität"

Nutzen	Inhaltliche Qualität	Sprachliche Qualität	Formale Qualität
Für Studium	0,12*	n.s.	n.s.
Für Berufsausbildung	0,12*	0,10*	n.s.
Für pers. Entwicklung	0,11*	n.s.	0,12*

* Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

5.9 Zwischenbilanz

Im Zuge der Analyse lassen sich auf der Grundlage des Ratings drei Skalen zur inhaltlichen, formalen und sprachlichen Qualität erhärten, die jeweils einen wichtigen Aspekt der Wissenschaftspropädeutik abbilden. Die Skalenmittelwerte liegen durchweg im positiven Bereich, so dass die Bewertung der Qualität gesamthaft auf das Prädikat "zufriedenstellend" lautet. Der Umstand, dass das Rating gleichwohl deutlich schwächer ausfällt als die schulische Benotung der Arbeiten, hat vor allem damit zu tun, dass in das pädagogische Urteil formative Aspekte mit eingehen, die bei einer summativen Bewertung der Projektberichte im Blickwinkel der Wissenschaftspropädeutik ausser Betracht bleiben müssen.

Die Skalen zur inhaltlichen, formalen und sprachlichen Qualität laden zwar jeweils auf einem eigenen Faktor, jedoch sind sie dimensional nicht vollkommen unabhängig voneinander. Das tritt zum einen darin zu Tage, dass diese Skalen untereinander positiv korrelieren, wie auch darin, dass sie jeweils positiv mit dem Globalurteil korrelieren, das den Gesamteindruck der Rater zur jeweiligen Arbeit wiedergibt (und das aufgrund des sechsstufigen Bewertungsschemas eben doch einer Schulnote ähnelt).

Hinsichtlich der institutionellen Rahmenbedingungen sprechen viele Indizien dafür, dass fokussierende Vorgaben bei der Themenfindung im Vergleich zur freien Themenwahl für die Qualität der Arbeiten förderlicher sind. Der bei der Hypothesenbildung unterstellte kausale Zusammenhang, dass die freie Themenwahl die intrinsische Motivation stärkt und daraus wiederum bessere Leistungen resultieren, hat sich als nicht richtig erwiesen. Die Strukturierung der Projektphasen durch Vorgaben erweist sich als die stärkere Trumpfkarte; wobei der Orientierungswert dieser Vorgaben noch dadurch verstärkt werden mag, dass sich die Bindewirkung unter den Absolventinnen und Absolventen erhöht und die betreuenden Lehrpersonen ihre Ressourcen konzentrierter einsetzen können.

Leitfäden und andere offizielle Papiere nehmen die Maturandinnen und Maturanden eher beiläufig zur Kenntnis; Transparenz ergibt sich für sie am ehesten auf der Basis interpersonalen Kontakts mit den betreuenden Lehrkräften. Das wahrgenommene Commitment der Lehrpersonen ist unzweifelhaft ein wichtiger Erfolgsfaktor, indes ist die Frequenz der Kontakte zwischen Lehrenden und Lernenden für das Gelingen der Projekte ebenso wie der Zeitraum der Bearbeitung wenig bedeutsam.

Positive Einschätzungen zum Nutzen für akademische und berufliche Qualifikationen sowie die persönliche Entwicklungen stammen überzufällig oft von Autorinnen und Autoren, deren Arbeiten von überdurchschnittlicher Qualität sind. Allerdings lässt sich auch hier von der Korrelation nicht auf die Kausalität schliessen: Die Nutzenerwartung kann die Qualität der Resultate beeinflussen, umgekehrt mag das gelungene Resultat aber auch ex post die Nutzenerwartung steigern.

6 Fazit

Die im Rahmen dieses Projekts vollzogene Evaluation von mehr als vierhundert Maturaarbeiten aus der Deutschschweiz, der Romandie und dem Tessin kommt insgesamt zu einem positiven Befund. Mehrheitlich ist die wissenschaftspropädeutische Qualität dieser Arbeiten als zufriedenstellend einzustufen. Gleichwohl fällt auf, dass das Rating in der Regel kritischer ausfällt als die Benotung durch die Lehrpersonen. Dabei ist zu berücksichtigen, dass der Fokus der pädagogischen Bewertung sich vom Fokus der vorliegenden Untersuchung in mehrfacher Hinsicht unterscheidet. So befinden die Lehrpersonen in der Regel nicht allein über das schriftliche Produkt, sondern auch über die mündliche Präsentation. Zudem musste sich die vorliegende Qualitätsanalyse auf eine summative Dimension beschränken, da sie nicht prozessbegleitend angelegt werden konnte, während die Lehrpersonen auch den Lernfortschritt der Schülerinnen und Schüler im Projektverlauf vor Augen haben. Darüber hinaus ist zu betonen, dass die im Rahmen dieser Untersuchung durchgeführte Evaluation der Maturaarbeiten sich weitgehend auf den akademischen Blickwinkel der Studierfähigkeit beschränkt, während die Lehrpersonen bei der Wahrnehmung ihres Bildungsauftrags eine Vielzahl anderer Aspekte in die Bewertung mit einfliessen lassen können.

Die Analyse erhärtet die Hypothese, dass die inhaltliche Qualität der Arbeiten mit ihrer formalen und ihrer sprachlichen Qualität korreliert: Schülerinnen und Schüler, die bei den inhaltlichen Gütekriterien gut abschneiden, erzielen in der Regel auch bei den anderen Gütekriterien gute Werte, hingegen sind bei inhaltlich schwachen Arbeiten auch überdurchschnittlich viele formale und sprachliche Mängel festzustellen. Die Vorgabe von Rahmenthemen wirkt sich positiv auf die Qualität der Maturaarbeiten aus; im Vergleich schneiden Maturaarbeiten, deren Thema völlig frei gewählt ist, etwas schwächer ab. Dies dürfte jedoch letztlich weniger auf die freie Themenwahl selbst als vielmehr auf die daraus resultierende grössere Komplexität der Betreuungsaufgaben zurückzuführen sein.

In der Auswertung zeigt sich, dass der interpersonale Kontakt zwischen Lehrenden und Lernenden ein Schlüsselfaktor für das Gelingen der Projekte ist. Dabei ist weniger die Art und Frequenz der Kontakte entscheidend, sondern vielmehr die wahrgenommene Intensität der Unterstützung bei der Bewältigung kritischer Phasen. Dies beginnt bei der Formulierung des Themas und erstreckt sich über die Klärung von Methodenfragen bis hin zum formalen Aufbau des Berichts. Ebenso wenig wie die Frequenz der Kontakte mit Lehrpersonen ist auch die Bearbeitungsdauer der Projekte für die Qualität der Maturaarbeiten bedeutsam.

Für die Wirksamkeit der Betreuung ist es vorteilhaft, wenn die Zuständigkeit nicht durch die Schule dekretiert, sondern den Schülerinnen und Schülern die Chance eröffnet wird, hier eine Wahl zu treffen, die auch durch den Aspekt der Sympathie beeinflusst ist. Neben den Lehrpersonen ist ferner das soziale Umfeld der Schülerinnen und Schüler von Gewicht. Erstaunlicherweise zeigen sich hier indes nur für Geschwister und Peers schwach signifikante Effekte. Der von den Maturandinnen und Maturanden deklarierte Einfluss der Eltern zeitigt hinsichtlich der Qualität der Maturaarbeiten keinen nachweisbaren Effekt. Hier ist indes zu betonen, dass zum Umfeld der Schülerinnen und Schüler nur wenige Angaben vorliegen und der latente Einfluss der Eltern auf die Bildungsbiographien völlig ausser Betracht bleiben musste.

Der Umstand, dass Gruppenarbeiten in der Regel sprachlich und inhaltlich besser sind als Einzelarbeiten, dürfte wesentlich auf die Binnenkommunikation im Projektteam zurückzuführen sein. Dieser Befund gehört insofern mit zum Einfluss der Peers auf die Qualität der Arbeiten. Dass hingegen bei der formalen Qualität kein Unterschied zwischen Einzel- und Gruppenarbeiten festzustellen ist, ist vor allem auf die institutionelle Wirksamkeit formaler Vorgaben im Rahmen der pädagogischen Betreuung zurückzuführen.

Wider Erwarten zeigt sich in der Evaluation, dass die Qualität der Maturaarbeiten nicht sonderlich stark durch das individuelle Interesse oder durch spezielle Vorkenntnisse beeinflusst ist, die sich die Maturandinnen und Maturanden ex post selbst zuschreiben. Zwischen der Selbstzuschreibung eines Zuwachses an kognitiver Kompetenz und der sprachlichen sowie inhaltlichen Qualität der Arbeiten gibt es einen schwach signifikanten positiven Zusammenhang. Weiter nicht verwunderlich ist, dass diese retrospektive Einschätzung deutlich stärker mit der Benotung der Maturaarbeiten zusammenhängt.

Die wahrgenommene Transparenz des Verfahrens, der Zielsetzungen und der Benotungsaspekte hat keine aussagekräftigen Auswirkungen auf die Qualität der Arbeiten. Schulische Leitfäden und andere Gestaltungshilfen werden von den Schülerinnen und Schülern eher beiläufig zur Kenntnis genommen, sind aber deshalb nicht bedeutungslos: Ihre Wirksamkeit scheint sich eher indirekt während der Betreuung der Projekte zu entfalten. Dem entspricht, dass Vorbereitungsveranstaltungen nur einen schwachen Effekt auf einzelne Qualitätsaspekte haben. Signifikant ist der Einfluss von speziellen Orientierungsveranstaltungen und von Erfahrungen im Rahmen des "Lernens am Projekt" (LAP). Je höher dessen Nutzen eingeschätzt wird, desto besser steht es um die inhaltliche Qualität der Arbeiten. Schülerinnen und Schüler, die das Fehlen einer vorbereitenden Veranstaltung bemängeln, schneiden bei der Qualität ihrer Arbeiten eher schwächer ab.

Die Schülerinnen und Schüler tendieren dahin, der Maturaarbeit einen intrinsischen Wert zuzusprechen; nur in vergleichsweise wenigen Fällen wird das Projekt im Rückblick als nutzlos erachtet. Mehr als die Hälfte erwartet einen positiven Lerneffekt im Hinblick auf ein Studium, deutlich weniger als die Hälfte im Hinblick auf eine Ausbildung. Die inhaltliche Qualität korreliert positiv mit den Einschätzungen der Schülerinnen und Schüler

zum Nutzen der Arbeiten für die weitere akademische oder berufliche Qualifikation wie auch mit dem Nutzen für die persönliche Entwicklung.

Zwischen natur-, sozial- und geisteswissenschaftlich ausgerichteten Maturaarbeiten gibt es keine gravierenden Qualitätsunterschiede. Auch beim Geschlechtervergleich ergeben sich keine bedeutsamen Differenzen hinsichtlich der Qualität der Arbeiten. Anders verhält es sich hingegen beim Regionalvergleich: Wiewohl der Befund insgesamt für alle Landesteile positiv auffällt, stechen doch beträchtliche sprachregionale Unterschiede ins Auge: Die Romandie schneidet im Vergleich mit der Deutschschweiz besser ab. Auch die Werte für die Arbeiten aus dem Tessin liegen mit Ausnahme der formalen Qualität über den Einschätzungen der Arbeiten aus der Deutschschweiz, fallen jedoch insgesamt etwas schwächer aus als die Einschätzungen zu den Arbeiten aus der Romandie. Diese Unterschiede konnten statistisch im Rahmen dieses Projekts nicht vollends aufgeklärt werden. Die Absicht, im grösseren Umfang bilinguale Rater einzusetzen, die in verschiedenen Wissenschaftsbereichen über die Grenzen der Sprachregionen hinweg Maturaarbeiten bewerten, hat sich aus organisatorischen Gründen nicht realisieren lassen. Das führt dazu, dass für den Vergleich zwischen der Romandie und der Deutschschweiz kein Eichmass zur Verfügung steht (die Daten für den Kanton Tessin sind aufgrund der kleinen Teilstichprobe nur bedingt mit denen aus den anderen Landesteilen vergleichbar). So ist nicht vollkommen auszuschliessen, dass trotz der Identität des Instrumentariums und trotz der Einheitlichkeit der Raterschulung unterschiedliche sprachregionale Bewertungskulturen wirksam werden, die in unterschiedlichen Anspruchsniveaus zum Ausdruck kommen. Tatsächlich gibt es aber eine Reihe von Indizien, die dafür sprechen, dass das bessere Abschneiden der Romandie im Vergleich zur Deutschschweiz nicht auf Idiosynkrasien, sondern auf institutionelle Faktoren zurückzuführen ist. Nicht zuletzt ist hier auch auf den Umstand zu verweisen, dass die Organisation der Maturaarbeit im Tessin deutlich mehr Parallelen zur Romandie als zur Deutschschweiz aufweist.

Von der Warte tertiärer Bildungsinstitutionen aus ergibt sich für die Maturaarbeit prima facie das Bild, dass die Maturandinnen und Maturanden in der Lage sein sollten, den elementaren Anforderungen eines Hochschulstudiums gerecht werden zu können: Sowohl bei der Strukturierung von Texten, bei der Verwendung von Zitaten und beim Anlegen von Bibliographien legen die Maturaarbeiten davon Zeugnis ab, dass die gängigen akademischen Regeln bereits im Gymnasium eingeübt werden. Dasselbe gilt für die Wiedergabe von elementaren theoretischen und empirischen Sachverhalten sowie ansatzweise auch für die Gewinnung von Informationen durch Recherchen oder durch Experimente, wiewohl bei den schulischen Projekten deskriptive und reproduktive Anteile deutlich gewichtiger sind als analytische Anteile.

Indes ist hier eine wichtige Einschränkung zu machen: Bei den Maturaarbeiten handelt es sich bildungsbiographisch um eine einzigartige Schlussetappe auf dem Weg zur Hochschulreife. Die Einzigartigkeit dieser Projekte, die Reichweite der institutionellen Vorkehrungen, die Intensität der pädagogischen Betreuung und die Unterstützung durch das soziale Umfeld bilden Kontextfaktoren, deren Zusammenwirken eine hohe Qualität des Outcome begünstigt. Wirken diese Faktoren im Rahmen des Studienalltags nicht mehr in dieser Bündelung zusammen, ist zu erwarten, dass viele der ersten Studienarbeiten von Anfangssemestern hinter dem Niveau der Maturaarbeiten zurückbleiben. Es war im Rahmen dieses Projekts nicht möglich, die Nachhaltigkeit der Lernprozesse zu untersuchen, welche die Schülerinnen und Schüler während der Maturaarbeit durchlaufen. Von daher lässt sich keine Prognose abgeben, in welchem Ausmass sie nach Abschluss des Gymnasiums in der Lage sind, die gesammelte Projekterfahrung in akademische oder berufliche Lernprozesse zu transferieren.

Obwohl der Beweis der Nachhaltigkeit bislang fehlt, weisen viele Indizien darauf hin, dass es sich bei der Maturaarbeit um eine sinnvolle und ertragreiche Lernform handelt. Sowohl die Ergebnisse des Ratings wie auch die Aussagen der Schülerinnen und Schüler sprechen dafür, den Einsatz der Projektmethode im Gymnasium beizubehalten und in moderater Form weiter auszubauen.

Literatur- und Quellenverzeichnis zu Kapitel VI

Quellenverzeichnis zu Kapitel VI

Aargau: Verordnung über die Promotion und die Maturität an den Mittelschulen (Maturitätsverordnung) vom 23. Juni 1999.

Bern: Direktionsverordnung über den gymnasialen Unterricht im 9. Schuljahr und den Unterricht an Maturitätsschulen (MaSDV) vom 3. Juli 1997.

Bern: Maturitätsschulverordnung (MaSV) vom 27. November 1996.

Basel-Landschaft: Verordnung über die Maturitätsprüfungen vom 5. Juli 2005.

Basel-Landschaft: Weisungen der Rektorenkonferenz zu den Maturaarbeiten vom 9. Mai 2000.

Basel-Stadt: Verordnung betreffend die Maturitätsprüfungen im Kanton Basel-Stadt vom 28. März 2000.

Basel-Stadt: Reglement für die Maturaarbeit an den Gymnasien im Kanton Basel-Stadt (Reglement für die Maturaarbeit) vom 2. Februar 2001.

Freiburg: Reglement vom 15. April 1998 über die Gymnasialausbildung (GAR).

Freiburg: Reglement vom 17. September 2001 über die Maturitätsprüfungen (MPR).

Freiburg: Richtlinien der Direktion für Erziehung, Kultur und Sport vom 21. Januar 2004 betreffend die Durchführung der Maturaarbeit.

Genf: Règlement relatif à la formation gymnasiale au collège de Genève du 14 octobre 1998.

Graubünden: Verordnung über das Gymnasium im Kanton Graubünden (GymVO) vom 6. Juli 1999.

Jura: Règlement sur la reconnaissance des certificats de maturité gymnasiale (RRM) du 16 janvier 1995.

Luzern: Reglement für die Maturitätsprüfungen im Kanton Luzern vom 27. Mai 1999.

Luzern: Die Maturaarbeit an den Luzerner Maturitätsschulen; Weisungen der Maturitätskommission vom 11. Januar 2000, revidiert am 4. Juli 2000.

Neuenburg: Règlement des études des lycées cantonaux (admission, promotion et examens) du 13 mai 1997.

Obwalden: Reglement über die Maturitätsprüfungen vom 22. April 1997.

Obwalden: Rahmenbestimmungen für die Maturaarbeit vom 22. April 2002.

St. Gallen: Maturitätsprüfungsreglement des Gymnasiums vom 24. Juni 1998.

Thurgau: Verordnung des Regierungsrates über die Maturitätsabteilung an den thurgauischen Kantonsschulen vom 6. Juli 1999.

Tessin: Regolamento degli studi liceali (del 24 giugno 1997).

Waadt: Réglement des gymnases (RGY) du 7 mai 1997.

Waadt: Loi sur l'enseignement secondaire supérieur (LESS) du 17 septembre 1985.

Wallis: Reglement über die Schulzeit am Gymnasium und die Maturitätsprüfungen vom 10. April 2002.

Wallis: Weisungen vom 10. Dezember 2004 betreffend die Maturaarbeit.

Zug: Reglement über die Maturitätsprüfungen an den kantonalen und privaten Gymnasien vom 17. März und 21. August 1997.

Zürich: Reglement für die Maturitätsprüfungen an den Gymnasien des Kantons Zürich vom 10. März 1998.

Literaturverzeichnis zu Kapitel VI

Arbeitsstelle für Hochschuldidaktik (2006): *Leistungsnachweise in modularisierten Studiengängen*. Dossier. Zürich: Universität Zürich, Arbeitsstelle für Hochschuldidaktik.

Bieri Buschor, Christine & Forrer, Esther (2005): Cool, kompetent und kein bisschen weise? Überfachliche Kompetenzen junger Erwachsener am Übergang zwischen Schule und Beruf. Zürich, Chur: Verlag Rüegger (=Wissenschaftliche Reihe, Bd. 18).

Binder, Hans-Martin (2003): Maturaarbeit - Engagement und Belastung. Ergebnisse einer Maturandenbefragung in Luzern. In *Neue Zürcher Zeitung*, Nr. 109, 13.05.2003, S. 69.

Binder, Hans-Martin & Feller-Länzlinger, Ruth (2004): *Maturitätsreform im Kanton Aargau. Schlussbericht der externen Evaluation*. Luzern: Interface, Institut für Politikstudien.

Binder, Hans-Martin & Feller-Länzlinger, Ruth (2003): Maturaarbeit und Maturaprüfung nach neuem Maturitätsanerkennungsreglement MAR. Auswertung der schriftlichen Befragung der AbsolventInnen der Matura 2002 im Rahmen der externen Evaluation der Gymnasialreform im Kanton Luzern. Luzern: Interface, Institut für Politikstudien.

Bonati, Peter (2007): Was Gymnasium und Standards voneinander lernen können. In Peter Labudde (Hrsg.), *Bildungsstandards am Gymnasium. Korsett oder Katalysator?* Bern: hep, S. 47-56.

Bonati, Peter (2006): Gymnasium wohin? - Bojen für die Weiterfahrt. In *Gymnasium Helveticum*, Nr. 3, S. 5-15.

Bonati, Peter & Hadorn, Rudolf (2007): Matura- und andere selbstständige Arbeiten betreuen. Ein Handbuch für Lehrpersonen und Dozierende. Bern: h.e.p.

Carolus, Thomas, (2006): Hinweise zur Anfertigung einer Studien- oder Diplomarbeit. Siegen: Universität Siegen, Fachbereich Maschinenbau, Institut für Fluid- und Thermodynamik. Online unter: http://www.uni-siegen.de/fb11/iftsm/aktuelles/anleitungwissarbeiten_3.pdf [15.09.2008]

Criblez, Lucien (2005): Das Gymnasium im Stress. In *amv-aktuell*, Sonderheft 2005/1, S. 4-12.

Darbellay, Vincent (2002): Travail de maturité: Une esquisse de bilan. In: Kollegiumsbote - Le Message du Collège, H. 1, S. 12–14.

Davaud, Clairette & Hexel, Dagmar (2003): *La nouvelle maturité gymnasiale. Orientations et intérêts des élèves de 4e*. Genève: Service de la Recherche en Education. Online unter: http://www.geneve.ch/sred/publications/docsred/2003/NouvMatuGym.pdf [19.12.2007].

Deutscher Bildungsrat (1970): Empfehlungen der Bildungskommission: Strukturplan für das deutsche Bildungswesen. Stuttgart: Klett.

Dreyer, Hans Peter (2005): Gymnasium und Universität: entfremdete Geschwister. In *amv-aktuell*, Sonderheft 2005/1, S. 22-27.

Dubs, Rolf (1996): Innovation und Qualität am Gymnasium. In Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK) (Hrsg.), *Von der "Mittelschule von Morgen" zur Maturitätsreform 1995*. Bern: EDK, S. 117-133.

Eberle, Franz (2007): Zum Fundament von Bildungsstandards an Schweizer Gymnasien. In Peter Labudde (Hrsg.), *Bildungsstandards am Gymnasium. Korsett oder Katalysator?* Bern: hep, S. 37-45.

Eberle, Franz (2005): EVAMAR Phase II – Eine Übersicht über das Evaluationsprojekt. *Gymnasium Helveticum*, Nr. 6, S. 5-11.

EDK [Schweizerische Konferenz der kantonalen Erziehungsdirektoren] (2007): *Stärkung der Naturwissenschaften und Aufwertung der Maturaarbeit in der gymnasialen Ausbildung.* Medienmitteilung vom 27. Juni 2007. Bern: EDK.

EDK [Schweizerische Konferenz der kantonalen Erziehungsdirektoren] (Hrsg.) (2000): La secondaire II à venir. Rapport final du groupe de project secondaire II. Die Sekundarstufe II hat Zukunft. Schlussbericht der Projektgruppe Sekundarstufe II. Bern: EDK.

EDK [Schweizerische Konferenz der kantonalen Erziehungsdirektoren] (1995): *Verordnung des Bundesrates/Reglement der EDK über die Anerkennung von gymnasialen Maturitätsausweisen (MAR)* vom 16. Januar/15. Februar 1995.

EDK [Schweizerische Konferenz der kantonalen Erziehungsdirektoren] (1994): Rahmenlehrplan für die Maturitätsschulen vom 9. Juni 1994. Empfehlung an die Kantone gemäss Art. 3 des Schulkonkordats vom 29. Oktober 1970. Mit Handreichung zur Umsetzung. Bern: EDK.

EDK [Schweizerische Konferenz der kantonalen Erziehungsdirektoren] (Hrsg.) (1972): Mittelschule von Morgen. *Jahrbuch der Schweizerischen Konferenz der kantonalen Erziehungsdirektoren*, Jg. 58.

Erdfelder, Edgar / Jochen Musch & Lutz Cüpper (2002): *Richtlinien zur Gestaltung von schriftlichen Referaten, Hausarbeiten und Praktikumsberichten.* Mannheim: Universität Mannheim, Lehrstuhl Psychologie III. Online unter: http://www.psychologie.uni-mannheim.de/psycho3/Sonstiges/Richtlinien.pdf [15.09.2008].

Fontolliet, Pierre-Gérard (1996): Une maturité de demain – Objectifs et réalités d'une réforme. In EDK [Schweizerische Konferenz der kantonalen Erziehungsdirektoren], *Von der «Mittelschule von morgen» zur Maturitätsreform 1995*. Bern: EDK, S. 47-60.

Frey, Karl (1993): *Die Projektmethode.* Weinheim: Beltz.

Furrer, Hans (2000): Ressourcen – Kompetenzen – Performanz. Kompetenzmanagement für Fachleute der Erwachsenenbildung. Luzern, Zürich: Akademie für Erwachsenenbildung AEB.

Grin, François; Alliata, Roberta; Dozio, Edo; Labudde, Peter; Maag Merki, Katharina & Pagnossin, Elisabetta (2004): Evaluation de la réforme de la maturité 1995 (EVAMAR). Document supplémentaire de la partie 3. Module 2. Objectifs pédagogiques transversaux. Genève: Service de la recherche en éducation (SRED).

Habel, Werner (1990): Wissenschaftspropädeutik. Untersuchungen zur gymnasialen Bildungstheorie des 19. und 20. Jahrhunderts. Köln, Wien: Böhlau.

Hadorn, Rudolf (2004): *Unterlagen zum Atelier 6: Maturaarbeit.* [Unveröff. Handout, verteilt im Rahmen der Kader-Tagung "Matura light" vom 25. August 2004 in Bern.]

Heldmann, Werner (1998): Studieren heute. Erwartungen der einzelnen Studienfächer an ihre Studienanfänger. (2., völlig neubearb. Aufl.) Bad Honnef: Bock.

Hentig, Hartmut von (1996): *Bildung. Ein Essay*. München, Wien: Hanser.

Hirschi, Hans (2002): Lehrpläne vermutlich noch zu ehrgeizig. In *Mitteilungsblatt Kanton Luzern*, Nr. 5, S. 29-30.

Huber, Ludwig (1998): Allgemeine Studierfähigkeit, basale Fähigkeiten, Grundbildung. Zur aktuellen Diskussion um die gymnasiale Oberstufe. In Rudolf Messner, Erhard Wicke & Dorit Bosse (Hrsg.), *Die Zukunft der gymnasialen Oberstufe. Beiträge zu ihrer Weiterentwicklung*. Weinheim, Basel: Beltz, S. 150-181.

Huber, Ludwig (1997): Fähigkeit zum Studieren - Bildung durch Wissenschaft? Zum Problem der Passung zwischen Gymnasialer Oberstufe und Hochschule. In Eckart Liebau, Wolfgang Mack & Christoph Scheilke (Hrsg.), *Das Gymnasium. Alltag, Reform, Geschichte, Theorie.* Weinheim, München: Juventa, S. 333-351.

Hupka, Sandra & Dellenbach, Myriam (2004): *Konsequenzen der Maturitätsreform 1995 für das Bestehen der Matura. Ergebnisse der Notenerhebung im EVAMAR-Projekt*. Bern: Erziehungsdirektion des Kantons Bern, Bildungsplanung und Evaluation.

Ingenkamp, Karlheinz (Hg.) (1989). Die Fragwürdigkeit der Zensurengebung: Texte und Untersuchungsberichte. Weinheim, Basel: Beltz.

Inglin, Oswald (2007): Macht die Maturaarbeit nicht kaputt! *Gymnasium Helveticum,* (1), S. 27-28.

Jacob, Walter (2004): Maturaarbeiten: Tut doch endlich etwas! In *Gymnasium Helveticum,* (5), S. 5-8.

Köller, Olaf (2007): Bildungsmonitoring und Vergleichsarbeiten in Deutschland. In Peter Labudde (Hrsg.), *Bildungsstandards am Gymnasium. Korsett oder Katalysator?* Bern: hep, S. 65-73.

Köller, Olaf; Watermann, Rainer; Trautwein, Ulrich & Lüdtke, Oliver (2004): Wege zur Hochschulreife in Baden-Württemberg - Erweiterung von Bildungswegen und Studiereignung: Die grundlegenden Fragestellungen in TOSCA. In Olaf Köller, Rainer Watermann, Ulrich Trautwein & Oliver Lüdtke (Hrsg.): *Wege zur Hochschulreife in Baden-Württemberg. TOSCA - Eine Untersuchung an allgemein bildenden und beruflichen Gymnasie*n. Opladen: Leske + Budrich, S. 113-119.

Klieme, Eckhard et al. (2003): *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise* (2. Auflage). Berlin: Bundesministerium für Bildung und Forschung (BMBF).

KMK [Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland] (2006): Vereinbarung zur Gestaltung der gymnasialen Oberstufe in der Sekundarstufe II. Beschluss der Kultusministerkonferenz vom 07.07.1972 i.d.F. vom 02.06.2006. Online unter: http://www.kmk.org/doc/publ/Vereinb-z-Gestalt-d-gymOb-i-d-SekII.pdf [27.02.2008].

KMK [Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland] (1978): Empfehlungen zur Arbeit in der gymnasialen Oberstufe. *Bildung und Erziehung, 31*, S. 561-574.

Konegen-Grenier, Christiane (2002): *Studierfähigkeit und Hochschulzugang.* Köln: Deutscher Instituts-Verlag.

Kruse, Otto (1999): Keine Angst vor dem leeren Blatt. Ohne Schreibblockaden durchs Studium. (7. Auflage.) Frankfurt und New York: Campus.

Kunz, Christin M. (2003): Gesellenstück gymnasialer Bildung. Maturaarbeit auf der zweiten Stufe gymnasialer Reform. *Neue Zürcher Zeitung*, 13. Mai 2003, S. 69.

Linke, Angelika (2006): Die Matura in den Zeiten von Bologna. In *Gymnasium Helveticum,* Nr. 4, S. 6-10.

Loos, Barbara & Popp, Susanne (1998): Praxis der gymnasialen Oberstufe. Varianten fachübergreifenden und fächerverbindenden Lernens und Arbeitens. In Rudolf Messner, Erhard Wicke & Dorit Bosse (Hrsg.), *Die Zukunft der gymnasialen Oberstufe. Beiträge zu ihrer Weiterentwicklung*. Weinheim, Basel: Beltz, S. 126-149.

Lorenzen, Klaus F. (2002): *Wissenschaftliche Anforderungen an Diplomarbeiten und Kriterien ihrer Beurteilung*. (3., völlig neu bearb. Ausg. 10.2.2002, zuerst ersch. am 8.1.1997). Online unter: http://bui.haw-hamburg.de/pers/klaus.lorenzen/ASP/wisskrit.pdf [04.12.2006].

Maag Merki, Katharina (2006): Teil III: Bilanz. In Katharina Maag Merki (Hrsg.): *Lernort Gymnasium. Individuelle Entwicklungsverläufe und Schulerfahrungen*. Bern: Haupt, S. 175-194.

Maag Merki, Katharina & Leutwyler, Bruno (2006): Die Fähigkeit zum selbst regulierten Lernen. In Katharina Maag Merki (Hrsg.), *Lernort Gymnasium. Individuelle Entwicklungsverläufe und Schulerfahrungen*. Bern: Haupt, S. 79-137.

Meylan, Jean-Pierre (1996): Die Erneuerung des Gymnasiums und die Anerkennung der Maturitäten. Stationen der Debatte 1968-1995. In Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK) (Hrsg.), *Von der "Mittelschule von Morgen" zur Maturitätsreform 1995*. Bern: EDK, S. 7-45.

National Academy of Science, 1996: *National Science Education Standards*. Online unter: http://www.nap.edu/readingroom/books/nses/2.html#dsl [07.02.2008]

Neeley, Stacia Dunn (2005): *Academic Literacy*. (Second Edition). New York: Pearson

Notter, Philipp & Arnold, Claudia (2003): Der Übergang ins Studium. Bericht zu einem Projekt der Konferenz der Schweizerischen Gymnasialrektoren (KSGR) und der Rektorenkonferenz der Schweizer Universitäten (CRUS). Bern: Bundesamt für Bildung und Wissenschaft (BBW) (=Schriftenreihe BBW, Bd. 5d).

Nussbaumer, Markus & Sieber, Peter (1994): Texte analysieren mit dem Zürcher Textanalyseraster. In Peter Sieber (Hrsg.), *Sprachfähigkeiten - besser als ihr Ruf und nötiger denn je! Ergebnisse und Folgerungen aus einem Forschungsprojekt*. Aarau, Frankfurt a.M., Salzburg: Sauerländer, S. 142-185.

Oelkers, Jürgen (1997): Geschichte und Nutzen der Projektmethode. In Dagmar Hänsel (Hrsg.), *Handbuch Projektunterricht* (S. 13-30). Weinheim, Basel: Beltz.

Oelkers, Jürgen (2007): *Die Qualität der Schweizer Gymnasien*. Bern: h.e.p. Verlag.

Pagnossin, Elisabetta; Alliata, Roberta & Dozio, Edo (2005): Le travail de maturité. In Erich Ramseier et al., (Hrsg.) Evaluation der Maturitätsreform 1995 (EVAMAR). Neue Fächerstruktur, Pädagogische Ziele, Schulentwicklung. Schlussbericht zur Phase 1. Bern: Staatssekretariat für Bildung und Forschung, S. 171-191.

Ramseier, Erich et al. (Hrsg.) (2005): Evaluation der Maturitätsreform 1995 (EVAMAR). Neue Fächerstrukturen, Pädagogische Ziele, Schulentwicklung. Schlussbericht zur Phase 1. Bern: Staatssekretariat für Bildung und Forschung.

Ramseier, Erich (2005): Stichprobenverfahren und Fehlerschätzung. Evaluation der Maturitätsreform 1995 (EVAMAR). Ergänzungsband – Teil 1. Bern: Erziehungsdirektion des Kantons Bern.

Ramseier, Erich et al. (o.J.): *Befragung Abschlussklassen*. Bern: Schweizerische Konferenz der kantonalen Erziehungsdirektoren und Bundesamt für Bildung und Wissenschaft. Online unter: http://www.sbf.admin.ch/evamar/FB/FB_EVAMAR_Abschluss_d.pdf [11.04.2008].

Ramseier, Erich; Allraum, Jürgen & Stalder, Ursula (2005): Neue Fächerstruktur und Ausbildungserfolg. In Ramseier et al., Evaluation der Maturitätsreform 1995 (EVAMAR). Neue Fächerstrukturen, Pädagogische Ziele, Schulentwicklung. Schlussbericht zur Phase 1. Bern: Staatssekretariat für Bildung und Forschung, S. 45-155.

Rheinberg, Falko (2002): Bezugsnormen und schulische Leistungsbeurteilung. In Franz E. Weinert (Hrsg.), *Leistungsmessungen in Schulen.* (2. Auflage). Weinheim, Basel: Beltz, S. 59-71.

Rychen, Dominique Simone & Salganik, Laura Hersh (2003): A holistic model of competence. In Dominique Simone Rychen & Laura Hersh Salganik (Hrsg.), *Key Competencies for a Successful Life and a Well-Functioning Society*. Göttingen: Hogrefe & Huber, S. 41-62.

Schäfer-Koch, Karin (1997): Studienvorbereitung. Zur Genese eines Problems. In Eckart Liebau et al. (Hrsg.), *Das Gymnasium. Alltag, Reform, Geschichte, Theorie*. Weinheim, München: Juventa, S. 189-201.

Schmid, Adrian (2002): Maturaarbeit: Last oder Lust. *Kollegiumsbote - Le Message du Collège*, (1), S. 10-11.

Schmidt, Arno (1994): Das Gymnasium im Aufwind. Entwicklung, Struktur, Probleme seiner Oberstufe. Aachen-Hahn: Hahner.

Schoch, Jürg (2003): Viel Aufwand für wenig Interdisziplinarität. Zur Reform der schweizerischen Gymnasien. *Neue Zürcher Zeitung*, 25. März 2003, S. 61.

Schräder-Naef, Regula D. (1980): Von der Mittelschule zur Hochschule. Ergebnisse einer Befragung zur Reform der gymnasialen Oberstufe, zur Studienwahl und zu Übertritts- und Studienproblemen. Bern, Stuttgart: Haupt.

Stocker, Eugen (2006): Regards sur le parcours gymnasial des jeune Vaudois. Incidences des nouvelles réglementations sur la réussite des élèves. Lausanne: Unité de recherché pour le pilotage des systèmes pédagogiques (URSP).

Tenorth, Heinz-Elmar (1994): *„Alle alles zu lehren". Möglichkeiten und Perspektive allgemeiner Bildung*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Trautwein, Ulrich, Lüdtke, Oliver & Husemann, Nicole (2006): Die Qualität der Studienvorbereitung in der gymnasialen Oberstufe: Eine Längsschnittstudie mit Absolventen von allgemeinbildenden und beruflichen Gymnasien. *Jahrbuch Jugendforschung, 6*, S. 47-67.

Trautwein, Ulrich & Lüdtke, Oliver (2004): Aspekte von Wissenschaftspropädeutik und Studierfähigkeit. In Olaf Köller, Rainer Watermann, Ulrich Trautwein & Oliver Lüdtke (Hrsg.): *Wege zur Hochschulreife in Baden-Württemberg. TOSCA - Eine Untersuchung an allgemein bildenden und beruflichen Gymnasie*n. Opladen: Leske + Budrich, S. 327-366.

Tschopp, Urs (2001): Zwischen Stühlen und Bänken - Lernen am Gymnasium zwischen Disziplin und Disziplinen. Versuch über die Interdsiziplinarität und ihre Stellung in der gymnasialen Bildung. In *Gymnasium Helveticum*, Nr. 5, S. 7-15.

USO [Union der Schülerorganisationen] (2006): Schülerorganisationen: Endlich 'zählt' die Maturaarbeit! *Medienmitteilung der USO vom 27.12.2006*.

Wallmeier, Martin (2007): Richtlinien zur Anfertigung wissenschaftlicher Arbeiten. Fribourg: Universität Fribourg, Departement für Betriebswirtschaftslehre, Lehrstuhl für Finanzmanagement und Rechnungswesen. Online unter: http://www.unifr.ch/finanzmanagement/lehre/Richtlinien_wiss_Arbeiten.pdf. [15.09.2008]

Weder, Hans (2005): Der Anspruch an die gymnasiale Bildung aus universitärer Sicht. In *amv-aktuell*, Sonderheft 2005/1, S. 17-18.

Weinert, Franz (2001): Concept of Competence: A Conceptual Clarification. In Dominique Rychen & Laura Salganik (Hrsg.), *Defining and Selecting Key Competencies*. Göttingen: Hogrefe & Huber, S. 45-65.

Weinert, Franz (2002):Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In ders. (Hrsg.), *Leistungsmessungen in Schulen* (2., unveränderte Aufl.). Göttingen: Hogrefe & Huber, S. 17-31.

Wilhelmer, Hermann (2005): Fach(bereichs)arbeiten an berufsbildenden Schulen. Besichtigung eines Modells am Beispiel der HLW und Modeschule Klagenfurt. *Informationen zur Deutschdidaktik (ide)*, (Heft 3), S. 84-88.

Winter, Felix (2006): Wie lernt man eine Maturaarbeit zu schreiben? Wissenschaftspropädeutisches Schreiben mit Diagnose und Rückmeldung. In: Otto Kruse, Katja Berger und Marianne Ulmi (Hrsg.): Prozessorientierte Schreibdidaktik. Schreibtraining für Schule, Studium und Beruf. Bern, Stuttgart, Wien: Haupt. S. 49–68.

Wolter, Andrä (1997): Das deutsche Gymnasium zwischen Quantität und Qualität. Die Entwicklung des Gymnasiums und der Wandel gesellschaftlichen Wissens. Oldenburg: Bibliotheks- und Informationssystem der Universität Oldenburg (bis).

Zillig, Elisabeth (2004): Gymnasiale Maturaarbeit und Interdisziplinäre Projektarbeit der Berufsmaturität. Ein Vergleich. Bern: Eidgenössische Berufsmaturitätskommission (EBMK).

Tabellen und Abbildungen zu Kapitel VI

Abbildungsverzeichnis

Abb. VI.1: Wirkungsmodell zur Genese der Maturaarbeiten 289
Abb. VI.2: Rahmenbedingungen im Wirkungsmodell 295
Abb. VI.3: Schülerinnen und Schüler im Wirkungsmodell 295
Abb. VI.4: Gymnasien und spezifische Vorbereitungsveranstaltungen 310
Abb. VI.5: Möglichkeiten der Themenwahl (N=30) 310
Abb. VI.6: Formale Vorgaben (Anzahl Nennungen) 311
Abb. VI.7: Varianz der Bearbeitungszeit in Monaten 312
Abb. VI.8: Gewichtung der Abschlussbeurteilung 314
Abb. VI.9: Zufriedenheit mit der Unterstützung durch die begleitende(n) Lehrperson(en) (Angaben in Prozent) 323
Abb. VI.10: Häufigkeit der Kontakte mit den Betreuungspersonen (Angaben in Prozent) 323
Abb. VI.11: Häufigkeit der individuellen Besprechungen mit den Betreuungspersonen (Angaben in Prozent) 324
Abb. VI.12: Zufriedenheit mit der Häufigkeit der Gespräche (Angaben in Prozent) 325
Abb. VI.13: Einschätzung der Fachkompetenzen (Angaben in Prozent) 327
Abb. VI.14: Einschätzungen von Teilnehmerinnen und Teilnehmern zum Nutzen vorbereitender Veranstaltungen (Angaben in Prozent) 330
Abb. VI.15: Institutionelle und individuelle Realisierungsbedingungen für die Maturaarbeit 331
Abb. VI.16: Einschätzungen zu Aufwand und Nutzen (Angaben in Prozent) 332
Abb. VI.17: Einschätzungen im Hinblick auf Qualifikation und persönliche Entwicklung 333

Tabellenverzeichnis

Tab. VI.1: Hypothesen zu übergreifenden Qualitätsaspekten 297
Tab. VI.2: Hypothesen zur Themenwahl 297
Tab. VI.3: Hypothesen zur Vorbereitung 298
Tab. VI.4: Hypothesen zu den Realisierungsbedingungen 298
Tab. VI.5: Hypothesen zur Betreuung 299
Tab. VI.6: Hypothesen zur Motivation und Selbstwahrnehmung 300
Tab. VI.7: Hypothesen zum intrinsischen und extrinsischen Wert der Arbeit 300
Tab. VI.8: Grundgesamtheit der Untersuchung (Gymnasiastinnen und Gymnasiasten des 1. gymnasialen Bildungsjahres im Schuljahr 2004/05) 301
Tab. VI.9: Grundgesamtheit und Stichprobe der Gymnasien 302
Tab. VI.10: Inhalte der Zielsetzungen, 2006 (N=30) 308
Tab. VI.11: Themenwahl nach übergreifendem Fachgebiet (Angaben in Prozent, Mehrfachantworten möglich) 318
Tab. VI.12: Wahl des Themas für die Maturarbeit nach Sprachregion (Angaben in Prozent) 319
Tab. VI.13: Motive für die Themenwahl (Angaben in Prozent, Mehr-fachantworten möglich) 320
Tab. VI.14: Benotung der Maturaarbeit (Angaben in Prozent) 321
Tab. VI.15: Auswahl der Betreuungsperson nach Sprachregion. (Angaben in Prozent, Mehrfachantworten möglich) 322
Tab. VI.16: Unterstützung durch die Lehrpersonen (Mehrfachnennungen möglich) 325
Tab. VI.17: Bewertung der Unterstützungsangebote (Mittelwerte, Standardabweichungen) 326
Tab. VI.18: Weitere Unterstützung (Mehrfachantworten möglich) 327

Tab. VI.19: Vorbereitende Veranstaltung (Mehrfachnennungen) 329
Tab. VI.20: Einschätzungen zum tatsächlichen individuellen Zeitaufwand 332
Tab. VI.21: Wirkungen der Maturaarbeit 334
Tab. VI.22: Skalen "Arbeitsorganisation" und "Kognitive Kompetenz" 335
Tab. VI.23: Korrelation von Nutzen- und Kompetenzaspekten (selbstattribuiert) 335
Tab. VI.24: Anteil der Ratingdifferenzen ≤ 2 Skalenpunkte (nach Kriterium, in Prozent) 337
Tab. VI.25: Zuordnung nach Fachgebieten 338
Tab. VI.26: Faktorladungen zu den Qualitätsmerkmalen der Maturaarbeiten 339
Tab. VI.27: Skala "Inhaltliche Qualität" 341
Tab. VI.28: Skala "Sprachliche Qualität" 341
Tab. VI.29: Skala "Formale Qualität" 341
Tab. VI.30: Korrelationen der drei Skalen "Inhaltliche Qualität", "Sprachliche Qualität" und "Formale Qualität" 342
Tab. VI.31: Mittelwerte und Standardabweichungen für die Kriterien der Skala "Inhaltliche Qualität" 343
Tab. VI.32: Mittelwerte und Standardabweichungen für die Kriterien der Skala "Sprachliche Qualität" 344
Tab. VI.33: Mittelwerte und Standardabweichungen für die Kriterien der Skala "Formale Qualität" 345
Tab. VI.34: Korrelation Globalurteil mit den drei Skalen "Inhaltliche Qualität", "Sprachliche Qualität" und "Formale Qualität" 346
Tab. VI.35: Korrelation der perzipierten Unterstützung durch die betreuenden Lehrperson mit den Qualitätsskalen 348
Tab. VI.36: Korrelation des eingeschätzten Nutzens mit den drei Skalen "Inhaltliche Qualität", "Sprachliche Qualität" und "Formale Qualität" 349

Anhang zu Kapitel VI

Anhang VI.A: Übersicht zu den Hypothesentests

Nr.	Hypothese	Bestätigung	
		ja	nein
1	**Übergreifende Qualitätsaspekte**		
1a	Es gibt eine positive Korrelation zwischen den inhaltlichen, formalen und sprachlichen Qualitätsaspekten der Maturaarbeiten.	**x**	
1b	Es gibt signifikante geschlechtsspezifische, sprachregionale und arbeitsformspezifische Unterschiede in Bezug auf die inhaltlichen, formalen und sprachlichen Qualitätsaspekte.		
	1b1: Geschlecht		x
	1b2: Region	**x**	
	1b3: Arbeitsform	**x**	
1c	Bei den einzelnen Qualitätsaspekten der Maturaarbeit gibt es keine Unterschiede aufgrund der disziplinären Zuordnung.	**(x)**	(x)
	1c1: Bei der inhaltlichen Qualität der Maturaarbeit gibt es keine signifikanten Unterschiede aufgrund der disziplinären Zuordnung.	**x**	
	1c2: Bei der formalen Qualität der Maturaarbeit gibt es keine signifikanten Unterschiede aufgrund der disziplinären Zuordnung.		x
	1c3: Bei der sprachlichen Qualität der Maturaarbeit gibt es keine signifikanten Unterschiede aufgrund der disziplinären Zuordnung.		x
1d	Interdisziplinär ausgerichtete Arbeiten weisen keine signifikanten qualitativen Unterschiede zu Arbeiten mit klarer disziplinärer Zuordnung auf.	**x**	
2a	**Themenwahl**		
	Je höher die Freiheit bei der Themenwahl, desto besser die inhaltliche Qualität der Maturaarbeit.		x
2b	**Vorbereitung**		
2b1	Intensive Vorbereitungsveranstaltungen (Intensiv-/Projektwoche, LaP. Einführungskurse, ev. Andere) wirken sich positiv auf die Qualität der Arbeit aus.	**(x)**	(x)
	2b1.1: Auf die inhaltliche Qualität	**x**	
	2b1.2: Auf die formale Qualität		x
	2b1.3: Auf die sprachliche Qualität	**x**	
2b2	Die Maturaarbeiten derjenigen Schülerinnen und Schüler, welche lediglich Informationsveranstaltungen oder gar keine vorbereitende Veranstaltungen besucht haben, weisen tendenziell eine geringere Qualität auf.	**x**	
2b3	Als nützlich wahrgenommene Vorbereitungsveranstaltungen wirken sich positiv auf die Qualität aus.	**x**	
2c	**Realisierungsbedingungen**		
2c1	Es gibt einen Zusammenhang zwischen der institutionell vorgegebenen Bearbeitungszeitspanne und der Qualität der Maturaarbeit.		x

Nr.	Hypothese	Bestätigung	
		ja	nein
2c2	Die Verfügbarkeit und Kenntnis eines Leitfadens führt zu einer höheren Qualität der Maturaarbeit.		x
2c3	Die wahrgenommene Transparenz der Abläufe wirkt sich positiv auf die Qualität der Arbeit aus.		x
2c4	Die wahrgenommene Transparenz der Bewertungskriterien wirkt sich positiv auf die Qualität der Arbeit aus.		x
2c5	Es gibt keinen Zusammenhang zwischen der Beurteilung der Bewertungskriterien und der Qualität.		x
3	**Betreuung**		
3a	Je besser die wahrgenommene Betreuung, desto besser die Qualität der Maturaarbeit.		x
3b	Die Zufriedenheit mit der Unterstützung bei der Formulierung der Fragestellung und der Unterstützung in Bezug auf den Inhalt der Arbeit korreliert positiv mit der inhaltlichen Qualität der Arbeit.		x
3c	Die Zufriedenheit mit der Unterstützung in Bezug auf formale Grundlagen sowie den Aufbau und die Gliederung der Arbeit korreliert positiv mit der formalen Qualität der Arbeit.	**(x)**	(x)
	3c1: formale Grundlagen	**x**	
	3c2: Aufbau und Gliederung		x
3d	Je höher die Freiheit bei der Betreuerwahl, desto positiver wird die Betreuung wahrgenommen.	**x**	
3e	Zwischen der Häufigkeit sowie der Form der Betreuungskontakte und der Qualität der Maturaarbeit gibt es keinen Zusammenhang.	**x**	
4a	**Motivation und Selbstwahrnehmung**		
4a1	Der wahrgenommene Zeitaufwand korreliert positiv mit der Qualität der Arbeiten.		x
4a2	Unterstützung durch das eigene Netzwerk (Eltern, Geschwister und Peers) wirkt sich positiv auf die Qualität der Maturaarbeit aus.	**(x)**	(x)
	4a2.1: Eltern		x
	4a2.2: Geschwister	**x**	
	4a2.3: Peers	**x**	
4a3	Vorkenntnisse zum Thema wirken sich positiv auf die inhaltliche Qualität aus.		x
4a4	Die Kompetenzwahrnehmung im Hinblick auf Vorkenntnisse und Selbstständigkeit korreliert positiv mit der Qualität.		x
4a5	Das retrospektiv deklarierte Interesse korreliert positiv mit der Qualität.		x
4b	**Intrinsischer und extrinsischer Wert der Arbeit**		
4b1	Die Qualität der Arbeit korreliert positiv mit der Einschätzung des Nutzens für die weitere Qualifikation.		x
4b2	Die Einschätzung des Nutzens korreliert positiv mit der Einschätzung des eigenen Kompetenzzuwachses.		x
4b3	Die Bewertung der Arbeit durch LP korreliert positiv mit der Einschätzung des eigenen Lernzuwachses.	**x**	

Anhang VI.B: Raster Rahmenbedingungen

Beispiel: Auswertungsbogen Kantonsschule X

Kriterium	Ausprägung / Bemerkungen	
Schulspezifische Rechtsgrundlagen	Ja	Reglement, Promotionsordnung, Weisung
Zuständigkeit Organisation	Ja	Koordinationsstelle (inkl. Schulleitungsmitglied)
Leitfaden	Ja	In: Weisung Themen: Rechtsgrundlagen, Ziele, Arbeitsorganisation (Einzel/Gruppen), Themenwahl, Betreuung, Umfang, Ehrenwort, Präsentation, Beurteilung, Rekurse, Termine, Organisation
Zielbestimmungen	Ja	In: Weisungen Themen: Eigenständigkeit, Schriftlichkeit, fachadäquate Methoden; mündliche Präsentation,
Vorbereitung	Ja	Infoveranstaltung
Themenwahl	Frei	Fach oder Mehrfach
Typen		(1) schriftlich, (2) schriftlich-kommentiert, (3) praktisch-gestalterisch
Einzel/Gruppe		Keine Bedingungen bekannt
Dokumentation Arbeitsprozess	Ja	Arbeitsjournal
Vorgaben Zeitaufwand	Nein	
Formale Vorgaben	Ja	Länge: 15-20 S exkl. Anhang; Empfehlung Format
		Gliederung Titel/Inhaltsverz./Vorwort - Einleitung - Hauptteil - Zusammenfassung - Quellen-Literaturverzeichnis
		Zitieranleitung
Arbeitsvertrag	Nein	
Unterrichtszeit	Ja	2 SWS
Vorbereitung Betreuungsperson	Nein	Keine Angaben zu Vorbereitung und Entlastung
Mandate	2/4	Pflicht/Maximum
Zeit pro Mandat	k.A.	Lehrperson muss regelmässig Besprechungen durchführen
Suchprimat	S	(S=Schüler, LP=Lehrer)
Mögliche Betreuende		Lehrperson & Lehrbeauftragte (intern)
Konsequenzen	FN, Z, ZP	(FN=Fachnote, Z=Zeugnis, ZP=Zulass. Präsentation)
Präsentation (Dauer, Examinatoren.)		15Min.(Prä) / 15Min (Diskussion), BetreuerIn & KoreferentIn
Gewichtung (Arbeit/Präs.)	66/33	Schriftliche Arbeit: Aufteilung nach Inhalt, Form und Arbeitsprozess
Rückmeldung		Mündlich, Note
Bewertung Gruppen		Teamnote oder n. Absprache auch Einzelbewertung
Bearbeitungsdauer		52 Wochen (Definitive Themenwahl bis Abgabe schriftliche Arbeit)

Anhang VI:C: Gewichtungsfaktoren

(1) Gewichtung nach Sprachregion

Sprachregion	Gewichtungsfaktor
Deutschschweiz	0.876325
Romandie	1.2
Tessin	1.170909

(2) Gewichtung nach Geschlecht

Geschlecht	Gewichtungsfaktor
weiblich	0.9533185
männlich	1.0670425

Anmerkung: Die Gewichtungen wurden beim Vergleich mit den Befunden aus EVAMAR I zum Zweck der besseren Vergleichbarkeit eingesetzt. Ansonsten beruhen die hier dargestellten Resultate durchweg auf ungewichteten Daten.

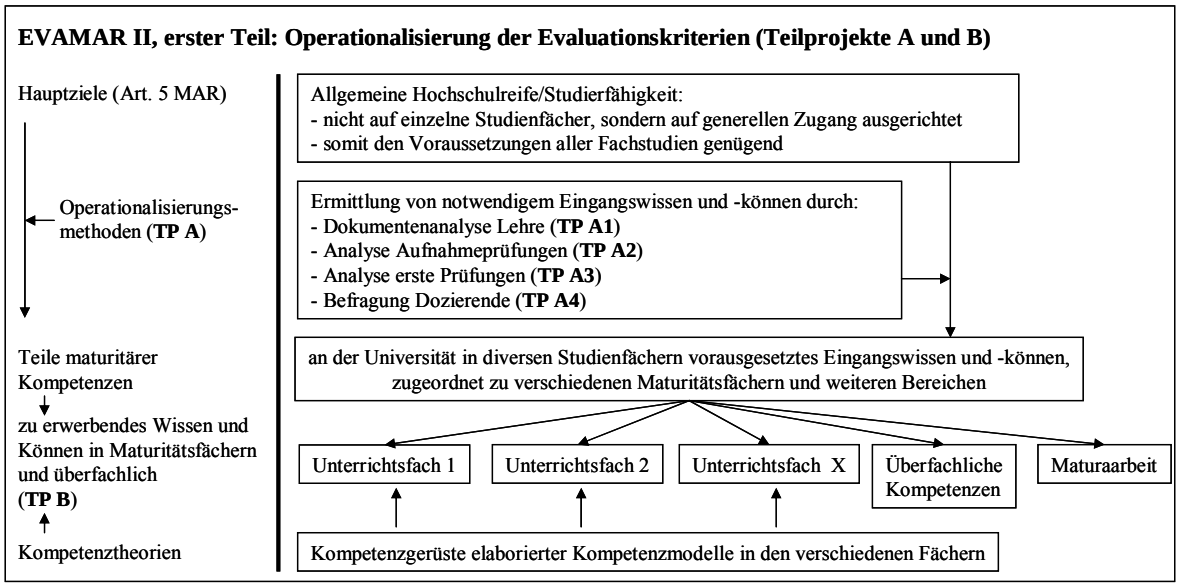
VII Zusammenfassung

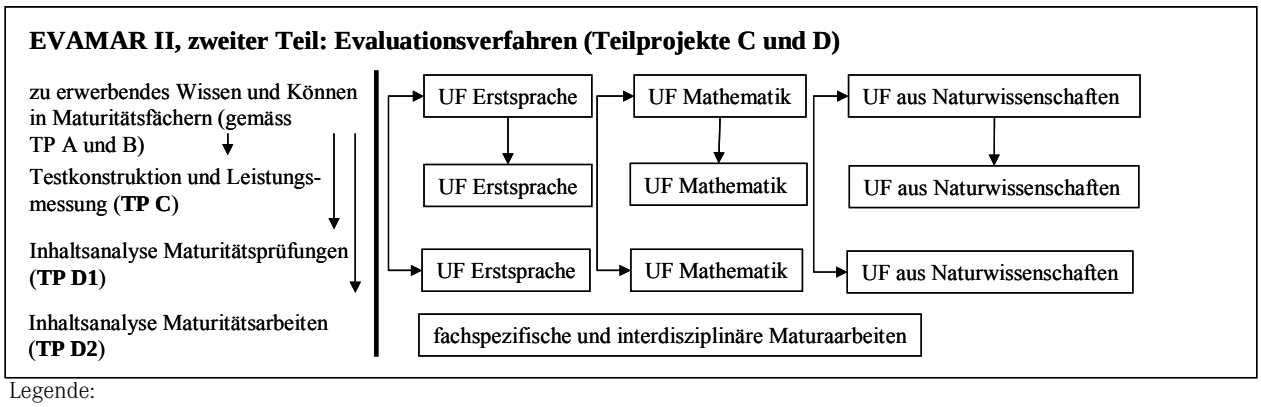
1 Fragestellungen

Eine evaluative Erfassung des Ausbildungsstandes muss sich sinnvollerweise an den Zielen der Ausbildung orientieren. Die Hauptziele der gymnasialen Bildung sind gemäss Art. 5 des MAR 95 (1995) die Erlangung „jener persönlichen Reife“, die erstens „Voraussetzung für ein Hochschulstudium ist“ (Hochschulreife oder Studierfähigkeit), und die zweitens „auf anspruchsvolle Aufgaben in der Gesellschaft vorbereitet“ (vertiefte Gesellschaftsreife durch breite Allgemeinbildung). Diese beiden Ziele haben teils gleichgerichtete, teils aber auch eigenständige und mitunter breite curriculare Auswirkungen. Eine vollständige Operationalisierung dieser Ziele in konkret zu erwerbende Kompetenzen der Gymnasiastinnen und Gymnasiasten in allen fachlichen und überfachlichen Lernbereichen und eine entsprechende flächendeckende Messung des Ausbildungsstandes, wie sie in einem ersten Konzept des Auftragnehmers vorgeschlagen wurde, wäre ausserordentlich aufwändig. Deshalb hat der Auftraggeber die im Rahmen von EVAMAR II vorzunehmenden Untersuchungen auf die in Abbildung 1 in ihrem Zusammenhang skizzierten Fragestellungen bzw. Teilprojekte eingeschränkt. Die Konkretisierung der Ziele der Maturitätsbildung sollte für ausgewählte Elemente des Hauptzieles der Studierfähigkeit erfolgen. Das zweite Hauptziel (vertiefte Gesellschaftsreife durch breite Allgemeinbildung) wurde damit nur teilweise in die durch den Auftraggeber gewählte Ziel-Operationalisierung einbezogen (im Überschneidungsbereich).

Auf der Grundlage der für EVAMAR II ausgewählten Untersuchungen können somit nur über *Teil*aspekte der Studierfähigkeit Aussagen gemacht werden. Das Ziel, „auf anspruchsvolle Aufgaben in der Gesellschaft vorzubereiten“, wurde zudem nicht direkt untersucht (siehe auch die im Abschnitt 7 beschriebenen Grenzen der Untersuchung). Diese Einschränkungen sind bei der Interpretation der Ergebnisse und bei der Ableitung von Massnahmen für die künftige Gestaltung des Gymnasiums zu beachten!

Abb. VII.1: Für EVAMAR II ausgewählte Untersuchungen





Legende:
- Die Buchstaben A-D entsprechen den Teilprojekten.
- UF: Abkürzung für „Unterrichtsfach“
- TP: Abkürzung für „Teilprojekt“

Die nachfolgende Darstellung der Ergebnisse fällt für die Teilprojekte A und B (Voruntersuchungen) knapper aus als für die Teilprojekte C und D, weil in dieser Kurzfassung des Berichts das Schwergewicht auf der Darlegung der Hauptergebnisse liegt.

2 Ergebnisse I: Voruntersuchung zur Konkretisierung des zu Messenden

Inhaltsanalyse von Studienmaterialien und Prüfungen

In den Teilprojekten A1 und A3 wurde eine Inhaltsanalyse von Lehrmaterialien und Prüfungen durchgeführt, um in den Lehrveranstaltungen des ersten und zweiten Semesters vorausgesetztes Wissen und Können zu ermitteln. Die Untersuchung erfasste die 16 im Studienjahr 2004/2005 meist gewählten Fächer, die von etwa 70% aller Immatrikulierten in der Schweiz studiert wurden. Es wurden für jedes Fach jeweils jene zwei oder drei Universitäten angefragt, an denen zusammen mehr als 50% aller Studierenden dieses Faches eingeschrieben waren. Obwohl nicht alle kontaktierten Dozierenden ihre Studienunterlagen und Prüfungen bereit stellten, umfassten die eingegangen Unterlagen alle ausgewählten 16 Fächer. In Bezug auf die Analyse der ersten Prüfungen im Rahmen von TP A3 ist einschränkend noch festzuhalten, dass nicht für jedes der Studienfächer Prüfungen zur Verfügung standen, weil in manchen Studiengängen aufgrund der Studienstruktur innerhalb der analysierten ersten beiden Semester keine Prüfungen stattfanden.

Die Analyse von Unterrichtsmaterialien und Prüfungen zielte auf die in Abbildung 2 dargestellten Zusammenhänge zwischen Maturitäts- und Studienfächern. Die resultierende, in eine Datenbank eingelesene Sammlung von Wissens- und Könnenselementen bildet somit nicht einfach den Rahmenlehrplan und die Fachlehrpläne an Maturitätsschulen ab. Die in EVAMAR II angewandte Analyse- und Codiermethodik im Teilprojekt A (und B) musste neu entwickelt werden.

Abb. VII.2: Zusammenhang zwischen Studienfächern und Unterrichtsfächern

Wissen und Können aus →	Mathematik	Erstsprache	Geographie	Physik	Biologie	usw.: Bereich Y
für Studienfach ↓						
Germanistik	GM	GE	GG	GP	GB	GY
Recht	JM	JE	JG	JP	JB	JY
Soziologie	SM	SE	SG	SP	SB	SY
Biologie	BM	BE	BG	BP	BB	BY
usw.: X	XM	XE	XG	XP	XB	XY
	↓	↓	↓	↓	↓	↓
auszubildende Studierkompetenzen in	**Mathematik**	**Erstsprache**	Geographie	Physik	**Biologie**	Fach Y oder überfachlich
	↑	↑			↑	
elaborierte Kompetenzmodelle aus Mathematik, Erstsprache und Biologie						

Legende: GM = erforderliches Wissen und Können aus dem Maturitätsfach Mathematik (M) für das Studienfach Germanistik (G) usw.

Insgesamt kann man aus den untersuchten Lehrtexten und Prüfungen über alle Studienfächer hinweg auf folgende Tendenzen hinsichtlich des vorausgesetzten und anzuwendenden Wissens und Könnens aus den verschiedenen Maturitätsfächern schliessen:

- Bei der Zuordnung zu verschiedenen Wissenskategorien überwiegen die Einträge zur Kategorie Faktenwissen deutlich, vorausgesetztes prozedurales oder metakognitives Wissen wurde dagegen eher selten gefunden. Diese Häufung gilt insbesondere für Eintragungen bei den sprachlichen Schulfächern und betrifft daher zunächst die drei Erstsprachen Deutsch, Französisch und Italienisch sowie auch Englisch. Viele Eintragungen sind aber auch in den Fächern Latein und Griechisch zu finden, wobei es sich zum grossen Teil um Doppeleintragungen zusätzlich zur Erstsprache handelt. Darin widerspiegelt sich der Kontext der alten Sprachen zur Begrifflichkeit in vielen Fachwissenschaften.
- Die Zuordnung der Einträge zu den einzelnen Schulfächern fällt sehr unterschiedlich aus und hängt erwartungsgemäss stark vom analysierten Studienfach ab. So häufen sich z. B. bei der Analyse des Studienfachs Biologie Einträge bei den Schulfächern Biologie und Chemie, oder die wirtschaftswissenschaftlichen Studiengänge bauen stark auf ein bereits überraschend umfangreich vorausgesetztes, wirtschaftliches Begriffsrepertoire auf.
- Griechisch- und Lateinkenntnisse können überwiegend als hilfreich, aber nicht unbedingt als notwendig angesehen werden, so etwa in den Studiengängen Rechtswissenschaften, Geschichte oder Pädagogik; eine profunde Kenntnis von Latein und Griechisch dürfte jedoch vor allem – wie bereits erwähnt – Vorteile für den Wortschatz insgesamt mit sich bringen, der sich wiederum über alle Studiengänge hinweg als sehr bedeutsam erweist. Dabei ist aber zu berücksichtigen, dass Fachtermini auch ohne ein entsprechendes Sprachenlernen verständlich sein bzw. gemacht werden können.
- Neben den wichtigen Erstsprachkenntnissen und einem umfangreichen Fremdwortschatz wird von den Studierenden zum Teil ein gutes Leseverständnis gefordert: Insbesondere in den geistes- und sozialwissenschaftlichen Fächern Geschichte, Soziologie und Pädagogik häufen sich Eintragungen, die auf das Arbeiten mit komplexen Texten verweisen und eine entsprechende Lesekompetenz erfordern.
- Hinsichtlich der naturwissenschaftlichen Kenntnisse lassen sich bei allen nicht naturwissenschaftlichen Studienfächern (in unserem Falle alle Studienfächer ausser Physik, Biologie und Medizin) kaum Unterschiede in der Zahl der ermittelten Wissenselemente aus den Bereichen der Physik und der Chemie feststellen. Hingegen ist der Rückgriff auf biologische Begriffe und Konzepte in den Geistes- und Sozialwissenschaften deutlich häufiger als auf solche aus Physik und Chemie.

- Auch aus dem überfachlichen Bereich des Wissenschaftlichen Arbeitens werden Kenntnisse vorausgesetzt. Daneben werden aber auch in einigen Studienfächern Veranstaltungen angeboten, die in das Wissenschaftliche Arbeiten einführen.
- Die Analyse der Aufgaben und Prüfungen in Bezug auf das vorausgesetzte Wissen zeigt, dass meist auf wenig Vorwissen referiert wird, das bereits vom Gymnasium mitgebracht werden muss; als für die Prüfungen erforderliche kognitive Fähigkeiten wurden überwiegend die Fähigkeitsstufen „Erinnern von Wissen", „Verstehen", „Anwenden" und „Analysieren" befunden. Fähigkeiten des „Evaluierens, Kreierens" oder des „Generierens und Planens" werden hingegen von den Studierenden eher wenig gefordert, zumindest in den ersten beiden Studiensemestern.

Die in den Teilprojekten A1 und A3 gewonnenen Inhalte sollten vor allem als wichtigste Grundlage für die im weiteren Projektverlauf zu testenden Schulfächer Erstsprache, Mathematik sowie Biologie dienen.

Befragung von Dozierenden

Ziel der Befragung von Dozierenden war es zu erfahren, welches fachliche und überfachliche Wissen und Können sie bei Studierenden des ersten und zweiten Semesters für ihre Veranstaltungen jeweils voraussetzen. Darüber hinaus wurden sie auch nach auffälligen Lücken im fachlichen und überfachlichen Wissen der Studienanfängerinnen und -anfänger gefragt. Angeschrieben wurden 320 Dozierende von Deutschschweizer Universitäten, deren Lehrveranstaltungen zuvor auch für die Analyse der Studienunterlagen ausgewählt worden waren. Die Rücklaufquote lag bei 44%. Die Hauptergebnisse, die von einigen Professoren der Romandie und des Tessins im Wesentlichen auch als für ihr Sprachgebiet zutreffend eingeschätzt wurden, sind nachfolgend aufgeführt.

Die Dozierenden wurden zunächst gebeten einzuschätzen, für wie wichtig sie fachspezifisches Wissen und Können aus den verschiedenen gymnasialen Maturafächern sowie in den Bereichen „Benutzerwissen Informatik" und „Informatikfachwissen" für das Verständnis und den erfolgreichen Besuch ihrer Veranstaltungen erachten. Eine Rangierung der Mittelwerte, gebildet aus einer vierstufigen Skala für die Antwortmöglichkeiten, ergibt die grösste Bedeutungszumessung für Deutsch, Englisch, Informatik-Benutzerwissen und Mathematik (in dieser Reihenfolge). Biologie und Physik befinden sich in dieser Reihung in der Mitte der Rangskala und Chemie am Beginn des letzten Drittels.

Schlüsselt man die Antworten auf diese Frage nach den einzelnen Studienfächern auf, verbleibt Deutsch bzw. Erstsprache fast durchwegs an erster Stelle und Englisch sowie Informatik-Benutzerwissen auf den vorderen Rängen. Bei den anderen Fächern ergeben sich teilweise erhebliche Unterschiede in der Rangierung. Mathematik wird in den sprach- und geisteswissenschaftlichen Fächern erwartungsgemäss als wesentlich weniger wichtig eingestuft, bleibt aber für viele Studienfächer sehr bedeutsam. An vorderer Stelle ist jeweils jenes Maturafach zu finden, das dem eigenen Studienfach entspricht.

Bei einer weiteren Frage mussten die Dozierenden konkretes vorausgesetztes Wissen und Können aus den verschiedenen Fächern aufführen. Hier zeigte sich, dass Abweichungen zum curricularen Verständnis des jeweiligen Faches am Gymnasium bestehen können. Wenn Dozierende von der grossen Bedeutung von Deutsch sprechen, meinen sie vorwiegend „Sprachbeherrschung", Grammatik, Syntax u. ä.. Eigentliche Kerninhalte der Germanistik wie Sprach- oder Literaturgeschichte, sprachwissenschaftliche Interpretations- und Textanalysemethoden, Grundlagen der Linguistik, Dramatheorie usw. werden nur von einigen der Germanisten und einem Anglisten erwartet. Es verbergen sich also hinter dem Begriff „Fach Deutsch" unterschiedliche Vorstellungen hinsichtlich der Fachinhalte. Wenn Deutsch allgemein als wichtigstes Maturafach eingeschätzt wird, ist somit hauptsächlich deutsche Sprachkompetenz gemeint. Ähnlich verhält es sich beim Fach Englisch, bei dem die Fähigkeit zur Bearbeitung englischsprachiger Fachliteratur (und nicht etwa literaturgeschichtliche Kompetenz) im Mittelpunkt steht.

Die nächste Frage war jene nach Lücken im zuvor als wichtig bezeichneten Wissen und Können bei Studienanfängerinnen und -anfängern. In fast allen Fächern wurden solche genannt, herausragend viele in Erstsprache und Mathematik. In Erstsprache sind es dabei wieder im Wesentlichen Komponenten der „Sprachbeherrschung", deren „Dürftigkeit" vor allem Nicht-Germanistik-Dozierende monierten.

Die Dozierenden sollten in einer weiteren Frage auf einer siebenstufigen Skala einschätzen, für wie wichtig sie überfachliche Kompetenzen für das Verständnis und den erfolgreichen Besuch ihrer Veranstaltungen erachten. Alle vorgelegten Kompetenzen wurden als sehr wichtig (nahe bei der Stufe 7) bis mindestens „mittelwichtig" (leicht über Stufe 4) bewertet, und zwar in folgender Reihenfolge: (1) Selbstständiges Lernen, (2) Verantwortung für eigenes Lernen und Arbeiten übernehmen, (3) Selbstständiges Arbeiten, (4) Kritisches Denken, (5) Problemlösefähigkeit, (6) Schriftliche Ausdrucksfähigkeit, (7) Zeit effizient einteilen und Prioritäten setzen, (8) Mit Belastungen umgehen, (9) Hörverstehen im Zusammenhang mit Vorlesungen, (10) Mündliche Ausdrucksfähigkeit, (11) Umfangreiche Prüfungen vorbereiten, (12) Entwickeln neuer Ideen, (13) In kurzer Zeit viel Lernstoff verarbeiten, (14) Recherchefähigkeit, (15) Im Team arbeiten, (16) Vor Publikum auftreten. Im Weiteren konnten die befragten Dozierenden angeben, bei welchen dieser überfachlichen Kompetenzen sie Defizite feststellten. Am häufigsten genannt wurden die Folgenden: Schriftliche Ausdrucksfähigkeit (42.4%), Kritisches Denken (35%), Selbstständiges Arbeiten (30%), Selbstständiges Lernen (26%), Mündliche Ausdrucksfähigkeit (21%), Verantwortung für eigenes Lernen und Arbeiten übernehmen (20.6%). Die wenigsten Nennungen erhielten: Vor Publikum auftreten (3%) und Im Team arbeiten (3%).

Die Befragungsergebnisse haben im Hinblick auf die Tests fast ausnahmslos keine Korrektur der Erkenntnisse aus den Analysen A1-A3 erforderlich gemacht, sind aber als eigenständige Ergebnisse interessant. Als Ausnahmen dazu sind festzuhalten: In Erstsprache wurde auch ein Bereich „Grammatikalische Kompetenzen/Orthografie" in den Test für Erstsprache übernommen. In einem ergänzenden Fragebogen wurde um eine Selbsteinschätzung wichtiger überfachlicher Kompetenzen gebeten. Und schliesslich stützen die Befragungsergebnisse die Wahl von Biologie als naturwissenschaftliches Fach (siehe weiter hinten).

Analyse bestehender Aufnahmeprüfungen

Mit dem Teilprojekt A2 sollten anhand einer Analyse von Eingangsprüfungen an Schweizer Universitäten ergänzende Hinweise zu den Kompetenzen ermittelt werden, die für ein universitäres Studium vorausgesetzt werden. Zu diesen Prüfungen gehören zunächst die Aufnahmeprüfungen, die an den meisten Universitäten als Qualifikationsmöglichkeit für den Studienzugang angeboten werden, wenn der Nachweis des Bestehens einer Schweizer Matura fehlt. Eine erste Übersicht hat ergeben, dass diese Prüfungen in der Regel an das Prüfungsprogramm der Schweizerischen Maturitätsprüfungen gekoppelt, also den Maturaprüfungen an den Gymnasien sehr ähnlich sind. Weil letztere aber Untersuchungsgegenstand des Teilprojekts D1 sind und deshalb die Gefahr von zirkulären Schlüssen bestanden hätte, musste diese Kategorie von Zugangsprüfungen von der weiteren Analyse ausgeschlossen werden. Als weitere Kategorie von Eingangsprüfungen bieten sich eigentliche Eignungstests an, die sich auf explizite, theoretisch begründete und möglichst auch empirisch validierte Studienerfolgsprognosekonzepte stützen. Es hat sich gezeigt, dass in der Schweiz der Eignungstest für das Medizinstudium (EMS) das einzige Instrument ist, das diesen Anforderungen genügt. Er wird für die auf die beschränkten Studienplätze ausgerichtete Zugangsselektion für das Medizinstudium an verschiedenen Universitäten eingesetzt und enthält Aufgaben, die überfachliche Kompetenzen erfassen (kein Wissenstest). Da der Einsatz solcher Aufgaben die Möglichkeit bot, in Ergänzung zu den in EVAMAR II vorgesehenen Fachtests weitere studienrelevante Kompetenzen zu erheben, entschied die Steuergruppe des Projekts EVAMAR II, im Rahmen des Teilprojekts C in der Deutschschweiz auch Aufgaben aus früheren Versionen des EMS vorzulegen. Dieser Testteil war im Auftragskonzept noch nicht vorgesehen.

3 Kompetenzmodellierung – TP B

In der Planung von EVAMAR II war zunächst vorgesehen, die Ergebnisse der Analysen der Teilprojekte A1-A4 direkt als Kompetenzmodelle zu verdichten. Dieses Vorhaben erwies sich aber schon bald aus verschiedenen Gründen als zu schwierig. Hauptsächlich deshalb, weil die in Projektteil A untersuchten Lehrtexte und Fachaufgaben zwar direkte Schlüsse auf jenes unabdingbare Wissen und Können zulassen, welches Voraussetzung für deren problemlose Bearbeitung ist, diese Informationen aber nicht ausreichend sind als Grundlagenelemente eines kohärenten, den aktuellen lerntheoretischen Erkenntnissen entsprechenden und empirisch abgesicherten Kompetenzmodells, wie es in der aktuellen theoretischen Diskussion um Bildungsstandards vertreten wird. Zudem decken die ermittelten Sinn- oder Wissenseinheiten und kognitiven Fähigkeiten die Anforderungen der Universitäten nicht flächendeckend ab. Es handelt sich nur um eine Stichprobe, die als ausschliessliche Basis eines vollständigen Kompetenzmodells für Studierfähigkeit noch zu eng wäre. Es erwies sich deshalb als notwendig, bereits bestehende Kompetenzmodelle in den Fächern auf ihre Tauglichkeit hinsichtlich der Erhebungsziele von EVAMAR II zu untersuchen, sie zu übernehmen oder sie weiterzuentwickeln. Bei diesen datengestützten Adaptionen ging es im Wesentlichen darum, die eher allgemein und formal gehaltenen Beschreibungen der Bereiche bestehender Kompetenzmodelle anhand des spezifischen Wissens und Könnens, das in Projekt A ermittelt wurde, im Hinblick auf die in EVAMAR II anvisierte Schnittstellenproblematik weiter zu konkretisieren. Die Modellierung erfolgte für die drei untersuchten Fächer Erstsprache, Mathematik und Biologie. Im Sinne der Abgrenzung von dem in der laufenden Diskussion bereits belegten Begriff bzw. Konstrukt des „Kompetenzmodells“ mit umfassendem Anspruch werden die Ergebnisse dieser Modellierung als „Kompetenzraster“ bezeichnet.

Für alle drei Testfächer entstanden auf diese Weise Kompetenzfelder, welche zum im Projekt A ermittelten Wissen und Können passen. Im Projektteil C wurden dann zu diesen Feldern Aufgaben konstruiert, welche sich auf Fachinhalte aus Projekt A beziehen. Alle Testaufgaben können also sowohl in der gewählten Kompetenzmodellierung als auch in den Lehrmaterialien an Universitäten verortet werden. Diese konkrete und konsequente „Schnittstellen-Verknüpfung“ der Tests bzw. der Leistungsmessung im Projekt EVAMAR II wurde noch nie in einer derartigen inhaltlichen Nähe vorgenommen, sie ist also neuartig. Die folgenden bereits bestehenden Kompetenzmodelle wurden als am bedeutsamsten für EVAMAR II befunden: der Gemeinsame Europäische Referenzrahmen für Sprachen (GER), die Modellierung der Kompetenzen für Mathematik in der dritten internationalen Mathematik- und Naturwissenschaftsstudie TIMSS III sowie die Einheitlichen Prüfungsanforderungen in der Abiturprüfung Biologie (EPA) der Kultusministerkonferenz (KMK) in Kombination mit den Kapiteleinteilungen von drei deutschsprachigen Standardlehrwerken der Biologie auf der Gymnasialstufe (Stufe SII). Es ist zu betonen, dass die Kompetenzraster aus der Anlage des Projekts heraus nicht notwendigerweise den gültigen Curricula der drei Testfächer entsprechen. Das gilt insbesondere auch für Erstsprache. Viele curriculare Inhalte dieses Faches, denen aus anderen Gründen als der Vorbereitung auf viele Studienfächer grosse Bedeutung zukommt, sind im Kompetenzraster und dem darauf basierenden Test nicht enthalten.

4 Ergebnisse II: Leistungsmessung – TP C

Stichprobe und eingesetzte Tests

In Rahmen des Teilprojekts C wurden Erhebungen bei Maturandinnen und Maturanden durchgeführt. Sie setzten sich aus Tests in den drei Fachbereichen Erstsprache, Mathematik und Biologie, einer Zusatzbefragung sowie einem überfachlichen Fähigkeitstest (UFT) zusammen. Die Notwendigkeit der Zusatzbefragung und des Einsatzes des UFT ergab sich erst aufgrund der oben beschriebenen Voruntersuchungen. Der Einbezug eines naturwissenschaftlichen Fachs anstatt – wie zuerst vorgeschlagen – Englisch wurde vom Auftraggeber gewünscht. Die konkrete Wahl aus Biologie, Chemie und Physik war zum Zeitpunkt der Auftragserteilung noch offen. Die Steuergruppe entschied sich vor allem deshalb für Biologie, weil dieses Fach im Vergleich zu den anderen beiden für die grössere Zahl von Studienfächern Grundlagen liefert (siehe auch die Ergebnisse der Be-

fragung Dozierender), und weil die Art der Testleistungen sich am ehesten von den bereits in Mathematik und im UFT gemessenen unterscheidet. In der Zusatzbefragung wurde in Ergänzung zum Projekt EVAMAR I auch nach der Selbsteinschätzung weiterer Kompetenzen im Rahmen der Studierfähigkeit gefragt. In der Deutschschweiz war dieser Fragebogen aufgrund der dort verfügbaren Befragungszeit noch umfassender. Der UFT enthält Items, die in anderen Untersuchungen gut mit dem Studienerfolg in naturwissenschaftlichen Fächern korreliert haben. Er setzt sich aus validierten Items von vier Dimensionen des Eignungstests für das Medizinstudium (EMS) zusammen, die einen besonders hohen Vorhersagewert für Prüfungen in universitären Fachstudien haben. Dieser Zusatztest wurde nur in der Deutschschweiz eingesetzt, weil in der Einschätzung der Projektleitung die Beanspruchung zusätzlicher, in der ursprünglich kommunizierten Planung nicht vorgesehener Testzeit in der Romandie und dem Tessin nicht mehr zumutbar war.

Als Grundlage für die Konstruktion von Aufgaben dienten die im Teilprojekt B erstellten Kompetenzraster, die in der Ergebnisdatenbank der Teilprojekte A1 und A3 gespeicherten Wissens- und Könnenselemente sowie die entsprechenden konkreten Lehrmaterialien. Dabei wurde eine möglichst repräsentative Verteilung auf die Kompetenzfelder einerseits und die ermittelten Wissens- und Könnenselemente andererseits angestrebt. Die Ergebnisse der Befragung von Dozierenden in Teilprojekt A4 erforderten – mit der bereits erwähnten Ausnahme in Erstsprache – keine Korrektur dieser Grundlage. Um die ganze Breite der Kompetenzraster einbeziehen zu können, wurden insgesamt Testaufgaben entwickelt, die von einem Schüler oder einer Schülerin in 720 Minuten gelöst werden könnten. Da für die Erhebungen nicht so viel Zeit zur Verfügung stand, konnten nicht jeder Schülerin bzw. jedem Schüler alle Aufgaben vorgelegt werden. Pro Test konnten lediglich 45 Minuten verfügbar gemacht werden. Für eine zuverlässige Schätzung der Fähigkeiten der Schülerinnen und Schüler, aber auch der Schwierigkeit der Aufgaben, wurde deshalb ein rotierendes Testdesign gewählt, bei dem die Aufgaben untereinander verlinkt sind (Multi-Matrix-Design, bei dem die Testhefte über gleiche Aufgabenblöcke miteinander verknüpft sind, sodass sich das für die einzelne Person ermittelte Wissen und Können nicht nur auf das von ihr bearbeitete Testheft stützen muss). Die Testhefte wurden zufällig auf die Schülerinnen und Schüler verteilt, sodass jede Aufgabe in jeder Klasse jeweils von ähnlich vielen Schülerinnen und Schülern bearbeitet wurde. Bei der Wahl der Aufgabenformen dominierten wie bei vergleichbaren grossen Untersuchungen (TIMSS, PISA) die Kriterien Auswertungsobjektivität und Auswertungsökonomie. Es wurden möglichst wenig Aufgaben zur ausführlichen offenen Beantwortung eingesetzt (am meisten in Erstsprache), weil diese zur Sicherung der Objektivität aufwändige Codierungsarbeiten nach sich gezogen hätten. Die Beschränkung auf die schriftliche Durchführungsform hatte auch eine Einschränkung der testbaren Kompetenzfelder in Erstsprache zur Folge (z. B. Verzicht auf Hör- und Sprechkompetenzen).

Mit der gewählten Kombination von Testinstrumenten konnte insgesamt eine Messung von im Hinblick auf die Studierfähigkeit vielfältigen Aspekten sichergestellt werden (fachliches Wissen, allgemeine Sprachkompetenzen, mathematische Kompetenzen, biologische und allgemeinere naturwissenschaftliche Kompetenzen).

Die Erhebungen wurden in einem über alle Schulen vergleichbaren Zeitraum durchgeführt. Massgebend für den Erhebungszeitpunkt war die Bedingung, dass die Maturandinnen und Maturanden maximal drei Wochen vor Ende des regulären Unterrichts vor den Maturaprüfungen stehen sollten. Wegen der terminlichen Unterschiede zwischen verschiedenen Schulen ergab sich daraus ein Erhebungszeitraum von Ende April bis Anfang Juli 2007.

Zur Grundpopulation gehören die Gymnasiastinnen und Gymnasiasten der Schweiz, die im Sommer 2007 die Maturitätsprüfungen abgelegt haben. Von der Population ausgeschlossen waren die Gymnasiastinnen und Gymnasiasten der Kantone Basel-Landschaft und Genf. Im Kanton Basel-Landschaft findet die Maturitätsprüfung jeweils erst im Herbst statt, weshalb die Vergleichbarkeit nicht gegeben war. Die Gymnasien des Kantons Genf erhoben schon früh grosse Einwände gegen die Untersuchung, und die meisten Schulen verweigerten die Übermittlung der Angaben, die Ende 2006 eingeholt und zur Bestimmung der Grundgesamtheit notwendig gewesen wären. Die Steuergruppe des Projekts EVAMAR II verzichtete in der Folge aus verschiedenen Gründen darauf, eine allfällige, von den zuständigen Behörden von oben verordnete Teilnahme zu erwirken. Eine zwangsweise Teilnahme hätte ein erhebliches Verfälschungspotenzial für die Validität der Ergebnisse mit sich gebracht. Um verschiedene Vergleiche von Subpopulationen durchführen zu können, wurde die Stichprobe stratifiziert. Nicht

alle gewünschten Klassen haben an den Erhebungen teilgenommen. Problematisch ist aber einzig die aus verschiedenen Gründen niedrig ausgefallene Beteiligungsquote der Tessiner Schulen. Deshalb konnten die Ergebnisse des Tessins in diesem Teilprojekt nicht in die Auswertungen einbezogen werden. Die Rücklaufquote ohne Tessin beträgt insgesamt 91% bei den Klassen und 85% bei den Personen. Die tiefere zweite Zahl erklärt sich aus Absenzen an den Erhebungstagen. Insgesamt lagen auswertbare Daten von rund 3'800 Personen vor.

Gesamtergebnisse

Im Mittel wurden die Aufgaben des Erstsprachtests zu etwas mehr als der Hälfte richtig gelöst (im Bereich „Grammatische Kompetenz und Orthografie" deutlich mehr). Die geschätzten Personenfähigkeiten liegen im Durchschnitt in allen Kompetenzbereichen auf einem Niveau, auf dem die Maturandinnen und Maturanden Aufgaben lösen konnten, deren Anforderungen über einem mittleren Schwierigkeitsgrad liegen. Die Streuung auf Ebene der einzelnen Personen ist beachtlich. Mit Ausnahme des Bereichs „Grammatische Kompetenz und Orthografie" wurde jeweils auch die maximal mögliche Punktzahl erreicht. Eine zusätzliche Auswertung der über die Klassen aggregierten Werte ergibt eine erhebliche Streuung auch zwischen den Klassenmittelwerten. Die beste Klasse hat im Mittel doppelt so viele Aufgaben richtig gelöst wie die leistungsschwächste.

Die Aufgaben des Mathematiktests wurden im Mittel von den Testpersonen zu etwas weniger als der Hälfte richtig gelöst. Insgesamt liegen die geschätzten Personenfähigkeiten im Durchschnitt in allen Kompetenzbereichen auf einem Niveau, auf dem die Maturandinnen und Maturanden Aufgaben lösen konnten, deren Anforderungen ganz leicht unter einem mittleren Schwierigkeitsgrad liegen. Die Streuung zwischen den einzelnen Personen ist wieder erheblich. In allen Bereichen wurde auch das Punktemaximum erreicht. Eine zusätzliche Auswertung der über die Klassen aggregierten Werte ergibt wieder eine bemerkenswerte Streuung auch zwischen den Klassenmittelwerten. Die beste Klasse hat also im Mittel beinahe dreimal so viele Aufgaben richtig gelöst wie die leistungsschwächste.

Da an vielen Gymnasien das Fach Biologie bereits ein bis zwei Jahre vor den Maturaprüfungen abgeschlossen wird, wurden die Resultate für zwei Gruppen („abgeschlossen ja", „abgeschlossen nein") berechnet. Im Mittel wurden die Testaufgaben in Biologie ebenfalls knapp zur Hälfte richtig gelöst, wobei die Gruppe mit zurückliegendem Fachabschluss noch schlechter, die andere Gruppe hingegen überdurchschnittlich abgeschlossen hat. Die Biologieaufgaben wurden von der ersten Gruppe auch im Vergleich zu Erstsprache und Mathematik eher schlechter gelöst. D. h., auch die geschätzten Personenfähigkeiten liegen im Durchschnitt unter einem mittleren Fähigkeitsmass. Vermutlich spielt hier der Vergessenseffekt eine Rolle, der aber auch weit höher hätte ausfallen können. Im Gesamttotal wurde die maximal mögliche Punktzahl von niemandem erreicht, wenn zum Teil auch nur knapp verfehlt. Die Streuung auf Ebene der einzelnen Personen ist wieder beachtlich. In jedem Bereich wurde jeweils auch die maximal mögliche Punktzahl erreicht. Die zusätzliche Auswertung der über die Klassen aggregierten Werte ergibt nochmals eine erhebliche Streuung auch zwischen den Klassenmittelwerten. Wieder hat die beste Klasse im Mittel beinahe dreimal so viele Aufgaben richtig gelöst wie die schlechteste.

Die Auswertung der Testergebnisse offenbart in allen Bereichen eine erstaunlich breite Streuung, vor allem angesichts der Tatsache, dass die Schülerinnen und Schüler kurz vor der Verleihung der für alle Studienfächer geltenden, universalen Qualifikation „Hochschulreife" standen. Es kann deshalb davon ausgegangen werden, dass nicht alle Maturandinnen und Maturanden in allen drei getesteten Fachbereichen über Kompetenzen verfügen, die den universitären Anforderungen aller Studienfächer entsprechen. Deshalb wurde zum Vergleich auch ein Blick auf die Mittelwerte und die Streuungen effektiv vergebener Maturanoten in den drei Testfächern geworfen.

Ausser in den schriftlichen Mathematikprüfungen wurde der minimal mögliche Wert (Note 1) nicht vergeben. Die relative Streuung fällt insgesamt kleiner aus als bei den Tests. Es fällt zudem auf, dass die Streuung in Mathematik erheblich grösser ist als in den beiden anderen Fächern. Ebenso, dass in den schriftlichen Mathematikprüfungen die Durchschnittsnote bei 4.0, also gerade noch bei „genügend" liegt. Eine Auswertung der über die Klassen aggregierten Zahlen ergibt folgende Extremwerte der Klassendurchschnitte: Minima: Erstsprache

schriftlich 3.6, Erstsprache Gesamtnote 3.97, Mathematik schriftlich 2.88, Mathematik Gesamtnote 3.73, Biologie Erfahrungsnote 3.85, Naturwissenschaften Gesamtnote 4.0; Maxima: Erstsprache schriftlich 5.11, Erstsprache Gesamtnote 5.17, Mathematik schriftlich 5.48, Mathematik Gesamtnote 5.44, Biologie Erfahrungsnote 5.28, Naturwissenschaften Gesamtnote 5.31.

Bei Notenbewertungen steht fest, dass Noten unter 4.0 als ungenügende Qualifikation gelten. Demgemäss können aus einer weiteren Auswertung der Notendaten die nachfolgend aufgeführten Prozentanteile von Maturandinnen und Maturanden als in einzelnen Bereichen von den Schulen selbst als ungenügend qualifiziert angegeben werden (Note 3.9 oder tiefer). Schriftliche Erstsprachprüfung: 19.6%, Gesamt-Maturanote in Erstsprache: 4.7%, Schriftliche Mathematikprüfung: 41.4%, Gesamt-Maturanote in Mathematik: 24.4%, Biologie-Erfahrungsnote: 5.6%, Gesamt-Maturanote in Naturwissenschaften: 5.6%. In der schriftlichen Maturaprüfung in Mathematik wurden also 41.4% der Schweizer Maturandinnen und Maturanden als ungenügend bewertet. 19.6% erzielten in der schriftlichen Erstsprachprüfung (überwiegend in Form eines Aufsatzes) ungenügende Noten. Die an der Maturaprüfung mittels schriftlicher Leistungsmessung festgestellten Kompetenzen in diesen Fächern wurden somit durch die Schulen selbst bzw. durch die entsprechenden Lehrpersonen beim durch diese Zahlen wiedergegebenen Anteil von Maturandinnen und Maturanden als ungenügend bewertet. Die Zahlen steigen zudem beim Übergang in den genügenden Bewertungsbereich (Note 4 oder tiefer) sprunghaft an. Vermutlich neigt man bei dieser Grenze zum Aufrunden, sodass die objektive Zahl der ungenügenden Qualifikationen eher noch höher ausfallen dürfte. Die oben aufgrund der Leistungstestergebnisse von EVAMAR II angebrachte Vermutung partiell ungenügender Kompetenzen bestätigt deshalb eigentlich – zumindest teilweise – nur eine auch von den Examinatorinnen und Examinatoren ausgeübte Bewertungspraxis. Einzig die genauen Zahlen fallen anders aus. Bei der Notengebung dürfte zudem auch eine gewisse Anpassung an die Leistungsfähigkeit von Klassen eine Rolle spielen.

Die Aufgaben des UFT wurden im Mittel in den Bereichen „Quantitative und formale Probleme“, „Textverständnis“ und „Naturwissenschaftliches Grundverständnis“ zu etwas weniger als der Hälfte richtig gelöst. Einzig im Bereich „Diagramme und Tabellen“ liegen die Werte etwas über der Hälfte. Insgesamt befinden sich die Werte leicht unter den Ergebnissen des vergleichbaren Eignungstests für das Medizinstudium (EMS). Der tiefere Wert lässt sich damit erklären, dass bei Weitem nicht alle Maturandinnen und Maturanden den EMS ablegen und die entsprechende freiwillige Vorselektion bereits zu besseren Ergebnissen führt. Diese Vergleichswerte deuten auch darauf hin, dass sich die Probanden offenbar angestrengt haben. Die Ergebnisse im UFT sind somit auch als ein wichtiger Indikator für eine überwiegend seriöse Bearbeitung aller Tests zu deuten. Die geschätzten Personenfähigkeiten liegen im Durchschnitt in allen Kompetenzbereichen auf einem Niveau, auf dem die Maturandinnen und Maturanden Aufgaben lösen können, deren Anforderungen leicht über einem mittleren Schwierigkeitsgrad liegen. Auffällig ist wieder die breite Streuung. In allen Bereichen wurde von einzelnen Maturandinnen und Maturanden auch das Punktemaximum erreicht. Eine zusätzliche Auswertung der über die Klassen aggregierten Werte ergibt die im Vergleich zu den anderen Tests grössten Differenzen zwischen den Mittelwerten der jeweils leistungsstärksten und leistungsschwächsten Klasse.

In die ergänzende Zusatzbefragung wurde die Selbsteinschätzung folgender weiterer überfachlicher Kompetenzen aufgenommen, die (auch gemäss Befragung der Dozierenden) für den Studienalltag wichtig sind: Zeitplanung, zeiteffiziente Prüfungsvorbereitung, Zeiteffizienz, Konzentrationsfähigkeit, vernetztes Denken, Diskussionsfähigkeit, Perspektivenübernahme. Die Maturandinnen und Maturanden sollten auf einer Skala von 1 (= sehr selten) bis 5 bzw. 4 (=sehr oft) angeben, inwiefern Aussagen, welche die genannten überfachlichen Kompetenzen thematisieren, auf sie selbst zutreffen. Die selbst berichteten überfachlichen Kompetenzen sind unterschiedlich stark ausgeprägt. Während die Maturandinnen und Maturanden ihrer Einschätzung zufolge „selten bis manchmal“ aktive Zeitplanung betreiben und auch angeben, sich nur „selten bis manchmal“ auf eine Sache konzentrieren zu können, haben sie offensichtlich wenig Schwierigkeiten, sich effizient auf Prüfungen vorzubereiten oder ihre Zeit effizient zu nutzen.

Vergleich nach Straten

In einem ersten Vergleich wurden Regionen gemäss folgenden Überlegungen zur Stratenbildung miteinander verglichen: Zürich ist grösster Kanton; „Deutschschweiz 3J." umfasst jene Deutschschweizer Region, in der bis zum Maturitätsabschluss nur drei Jahre im eigentlichen Gymnasium absolviert werden müssen (BEd); „Deutschschweiz klein" umfasst die Kantone mit weniger als 15 Maturaklassen (AI, NW, OW, GL, UR, AR, SH, VSd, ZG); „Deutschschweiz gross" enthält die Kantone mit mehr als 15 Klassen (SZ, SO, TG, GR, BS, AG, SG, LU); „Romandie 1 (3J.)" besteht aus jenen Westschweizer Kantonen, in denen ebenfalls nur drei Jahre am eigentlichen Gymnasium absolviert werden müssen (BEf, JU, NE, VD); und „Romandie 2" umfasst die Westschweizer Kantone mit mindestens vier Jahren Ausbildungsdauer am eigentlichen Gymnasium (FR, VSf).

In Erstsprache schnitten die Gymnasien des Stratums „Deutschschweiz 3J." im Mittel signifikant am schlechtesten ab. Die Effektstärken zur jeweils besten Gruppe, Romandie 2, liegen zwischen gering und mittel. Auf der Ebene der einzelnen Kompetenzbereiche ergeben sich die folgenden Differenzierungen: Beim „Allgemeinen Leseverstehen" schnitten nur noch die kleinen Deutschschweizer Kantone und die Gruppe „Romandie 2" besser ab als die Straten mit den tiefsten Werten. Beim „Detaillierten Leseverstehen" ergibt sich eine nochmals differenziertere Gruppierung. Hier bildet auch das Stratum „Romandie 1 (3J.)" eine eigene Gruppe am unteren Ende. Im Bereich „Sprachreflexion/Wortschatz" fallen auch die grossen Deutschschweizer Kantone in die unterste Gruppe. Die Verteilung im Bereich „Grammatik" bestätigt im Wesentlichen das Bild des Gesamtergebnisses. Allerdings gehören hier auch die kleinen Deutschschweizer Kantone zur Gruppe der Besten. Insgesamt schnitten die Gymnasien mit nur dreijähriger Mindest-Dauer zumindest innerhalb des gleichen Sprachbereichs eher schlechter ab.

In Mathematik schnitten die Gymnasien mit nur dreijähriger Mindest-Dauer im Gesamtergebnis klar schlechter ab als jene mit mindestens vier Jahren, insbesondere in der Romandie. Die Effektstärken zwischen den Straten mit dem tiefsten („Romandie 1 (3J.)") und dem höchsten Wert (Romandie 2) sind mittlere. Dieses Bild wird insbesondere in den Bereichen „Stochastik" und „Verschiedenes" mit noch leicht höheren Effektstärken akzentuiert und umgekehrt im Bereich „Geometrie" etwas entschärft. Es scheint sich zu zeigen, dass die Schuldauer im Gymnasium in Mathematik eine noch grössere positive Auswirkung auf den Leistungsstand hat als in Erstsprache.

In Biologie liegt Zürich klar an der Spitze, mit einer hohen Effektstärke im Vergleich zum Stratum mit dem tiefsten Wert. Am unteren Ende gruppieren sich die Schulen aus der Romandie. Die Unterschiede zwischen den Gymnasien mit drei- und vierjähriger Mindest-Dauer am Schulort Gymnasium sind innerhalb der Romandie unbedeutend, in der Deutschschweiz hingegen nach wie vor vorhanden, und zwar zugunsten einer vierjährigen Mindest-Schuldauer.

Auch im überfachlichen Fähigkeitstest schneidet die Gruppe der Gymnasien mit dreijähriger Mindest-Dauer in allen Unterbereichen signifikant schlechter ab. Die Effektstärken zwischen den Extremgruppen liegen allerdings auf einem tiefen Niveau. Vermutlich bestätigt sich hier der Anspruch des Tests, Fähigkeiten zu prüfen, die relativ unabhängig vom Umfang des in den verschiedenen Fächern erarbeiteten Wissens sind. Die Schuldauer am Gymnasium scheint aber trotzdem bedeutsam zu bleiben, wenn auch in geringerem Ausmass.

Die in den Vergleichen der Fachtestergebnisse zwischen den Straten festgestellten mittleren bis grossen Unterschiede werfen die Frage auf, ob sich solche auch in den effektiv vergebenen Maturanoten in den drei Fächern zeigen. Die Maturanoten in Erstsprache unterscheiden sich zwischen den Straten ebenfalls signifikant, die Effektstärken zwischen den Extremwerten sind mittelgross. Die Rangierung ist aber im Vergleich zu jener bei den Testergebnissen in Erstsprache eine andere. Eine plausible Erklärung findet sich nicht so leicht. Sicher gehört zu den möglichen Gründen, dass ein gymnasiales Erstsprache-Curriculum weit mehr umfasst und z. T. anderes beinhaltet als mit den Tests gemessen wurde. Es ist augenfällig, dass in der Westschweiz über die Sprachgrenzen hinweg in Erstsprache tiefere Noten vergeben wurden als in der übrigen Schweiz. Die Differenz bei der Maturanote in Mathematik zwischen dem tiefsten und höchsten Mittelwert ist zwar signifikant, beträgt aber nur eine Zehntelnote. Die Effektstärke ist hier sehr gering. Interessant ist die im Vergleich zu den Testergebnissen

umgekehrte Reihenfolge der besten und schlechtesten Straten bei der Erfahrungsnote Biologie mit immerhin einer beinahe mittleren Effektstärke. Auch bei der Maturanote in Naturwissenschaften ergeben sich zwischen den Straten nur geringfügige Unterschiede. Insgesamt kann festgestellt werden, dass sich bei den Maturanoten keine oder nur noch kleine Unterschiede zwischen den Straten finden lassen. Vermutlich lässt sich dieser Befund damit erklären, dass Bewertungen an der Matura nicht an einer gesamtschweizerischen Norm erfolgen (können).

Vergleiche nach Schwerpunktfächern

Maturandinnen und Maturanden mit dem Schwerpunktfach „Alte Sprachen“ (vorwiegend Latein) haben im Erstsprachtest in allen Bereichen durchschnittlich am besten, jene mit dem Schwerpunktfach „Bildnerisches Gestalten“ ebenfalls in allen Bereichen und teilweise mit den Gruppen „Musik“ sowie „Wirtschaft und Recht“ am schlechtesten abgeschlossen. Die Effektstärken zwischen den Extremgruppen sind mittel. Es fällt auf, dass sich die Schülerinnen und Schüler des Sprachen-Schwerpunkts „Moderne Sprachen“ in einem Sprachtest „nur“ in der Mitte platzierten.

Maturandinnen und Maturanden des Schwerpunkts „Physik und Anwendungen der Mathematik“ haben in allen Bereichen des Mathematiktests klar am besten, die Gruppe „Bildnerisches Gestalten“ ebenfalls in allen Bereichen und teilweise zusammen mit der Gruppe „Philosophie/Pädagogik/Psychologie“ und dem Schwerpunkt „Musik“, am schlechtesten abgeschlossen. Das Resultat der besten Gruppe fällt erwartungsgemäss aus. Die Effektstärke zwischen dieser und der schlechtesten Gruppe ist aber ausserordentlich gross. Die Schülerinnen und Schüler des Schwerpunkts „Alte Sprache“ haben am zweitbesten abgeschnitten, noch vor jenen des Schwerpunkts „Biologie und Chemie“ und jenen des Schwerpunkts „Wirtschaft und Recht“. Die Effektstärken der zweitbesten Gruppe „Alte Sprachen“ zur jeweils schlechtesten liegen immer noch bei erheblichen Werten.

Im Biologietest haben die Maturandinnen und Maturanden des Schwerpunkts „Biologie und Chemie“ wiederum erwartungskonform in allen Bereichen die besten Ergebnisse erzielt. Auch in diesem Test liegen die Schülerinnen und Schüler des Schwerpunkts „Alte Sprachen“ in allen Bereichen an zweiter Stelle, noch vor jenen des anderen naturwissenschaftlichen Schwerpunkts „Physik und Anwendungen der Mathematik“. Die Ergebnisse der Gruppe des Schwerpunkts „Musik“ sind im Gesamtergebnis klar, in den Teilbereichen teilweise zusammen mit den Gruppen „Moderne Sprachen“, „Philosophie/Pädagogik/Psychologie“ sowie „Wirtschaft und Recht“ am schlechtesten ausgefallen. Die Effektstärken der schlechtesten Gruppe zur „Spezialistengruppe“ sind sehr hoch und bleiben im Vergleich zur zweitbesten Gruppe „Alte Sprachen“ immer noch moderat.

Im UFT ist die Reihenfolge der Ergebnisse in allen Unterbereichen dieselbe. Die Unterschiede zwischen der besten Gruppe („Physik und Anwendungen der Mathematik“) sowie der schlechtesten („Philosophie/Pädagogik/Psychologie“) sind erheblich (sehr grosse Effektstärke). Auch die Maturandinnen und Maturanden des Schwerpunkts „Bildnerisches Gestalten“ befinden sich in der Signifikanzgruppe mit den schlechtesten Ergebnissen.

Nachdem in den Vergleichen der Testergebnisse zwischen den Schwerpunktfach-Gruppen teilweise grosse Unterschiede ausgemacht werden können, interessiert, ob sich solche auch in den effektiv vergebenen Maturanoten in den drei Fächern zeigen. Die Effektstärke zwischen den Extremgruppen bleibt beim Vergleich der Maturanote in Erstsprache eine mittlere. Hervorzuheben sind die vergleichsweise zur Testrangierung verbesserten Bewertungen für die Schülerinnen und Schüler des Schwerpunkts „Musik“ und das Abrutschen der Maturandinnen und Maturanden des Schwerpunkts „Biologie und Chemie“. Beim Vergleich der Maturanote in Mathematik halbiert sich der mit der Effektstärke gemessene Unterschied zwischen den beiden Extremgruppen gegenüber dem Testunterschied beinahe, er bleibt aber sehr hoch. In diesem Fach stimmt auch die sonstige Rangierung recht gut mit jener der Testergebnisse überein. Das deutet darauf hin, dass es den bewertenden Personen in Mathema-

tik eher gelingt, sich bei der Benotung implizit an objektiven Güte-Standards zu orientieren. Beim Vergleich der Matura-Erfahrungsnote in Biologie fällt auf, dass die Effektstärke beim Unterschied zwischen den Extremgruppen auf ein mittleres Mass sinkt. Die Rangierung stimmt mit kleinen Abweichungen erstaunlich gut mit der Rangierung des Biologietests überein. Die grösste Abweichung kann für die Schülerinnen und Schüler des Schwerpunkts „Biologie und Chemie“ sowie „Bildnerisches Gestalten“ festgestellt werden. Erstere wurden mittels Noten vergleichsweise schlechter bewertet (und diesbezüglich gar von den „Altsprachlern“ überholt), letztere wurden besser bewertet.

Ergänzend zu diesen Mittelwertvergleichen wurden die Prozentzahlen für ungenügende Ergebnisse zusammengetragen. In Mathematik zeigen sich ungenügende Bewertungen in der schriftlichen Maturitätsprüfung für die Hälfte aller Maturandinnen oder mehr in den nach MAR 95 neuen Schwerpunkten „Philosophie/Pädagogik/Psychologie“, „Bildnerisches Gestalten“ und „Musik“. Im Weiteren haben offenbar Erfahrungsnoten kompensierende Wirkung für die in schriftlichen Prüfungen eher hohe Zahl an ungenügenden Bewertungen.

Vergleiche nach Geschlecht

Wie bei den meisten Untersuchungen wurden die Testergebnisse und die Maturanoten auch auf Unterschiede zwischen den Geschlechtern untersucht. Über alle Vergleiche hinweg ergaben sich mehrheitlich bessere Resultate für die Maturandinnen in Erstsprache (bei den Testergebnissen nur minim), hingegen mehrheitlich schlechtere in Mathematik, Biologie und im überfachlichen Fähigkeitstest. Erwähnenswert ist der Gegensatz zwischen dem signifikant schlechteren Abschneiden der Maturandinnen in allen Biologie-Testbereichen und der signifikant besseren Erfahrungsnote in Biologie für diese Gruppe. Es zeigte sich auch, dass Maturandinnen offenbar die schlechteren Noten in Physik mit besseren Noten in Biologie kompensierten (die Unterschiede bei der Erfahrungsnote in Chemie sind unbedeutend).

Vergleiche nach Maturitätsquoten

Die Maturitätsquoten in der Schweiz unterscheiden sich beträchtlich. Im Maturajahrgang 2007 haben gemäss Bundesstatistik im Kanton Tessin 29.5% der jungen Erwachsenen einen gymnasialen Abschluss erworben, im Kanton Obwalden waren es lediglich 11.2%. Diese Unterschiede müssen nicht zwingend eine entsprechende Chancenungleichheit widerspiegeln. Es ist aber von Interesse, ob an den Maturaprüfungen überall die gleichen Anforderungen gestellt werden. Im Rahmen von EVAMAR II lassen sich Teile der Kompetenzen der Maturandinnen und Maturanden aus Kantonen mit unterschiedlichen Maturitätsquoten miteinander vergleichen. Weil die Stichprobe nicht für jeden Kanton repräsentativ ist, wird nur eine Grobeinteilung der Quoten zu Gruppen vorgenommen. Überlagerungen mit anderen Effekten bleiben dabei offen, ebenso die Frage der Gleichheit von Anforderungen bei der Eingangsselektion der Gymnasien. Bei der Beurteilung der Chancengleichheit müsste nämlich auch letztere berücksichtigt werden, weil ein „Aufholen“ von Kompetenzdefiziten während der Gymnasialzeit durchaus möglich ist, sodass sich dann Eingangsunterschiede am Schluss nicht mehr zeigen, die Chancenungleichheit aber bleibt. Wegen der sich überlagernden Effekte wurde ein Vergleich zwischen Gruppen in einem oberen und einem unteren Bereich vorgenommen. Die Maturandinnen und Maturanden aus der Gruppe mit der tieferen Maturitätsquote erzielten in beinahe allen Test- und Notenbereichen die besseren Resultate. Die grössten Unterschiede lassen sich bei den Ergebnissen in Biologie finden (mindestens mittlerer Effekt). Bei den meisten Maturanoten fallen die Unterschiede in der gleichen Richtung aus wie bei den Testergebnissen, aber in geringerem Ausmass. Ausnahme davon sind die Erstsprachennoten, deren Differenzen die Unterschiede in den Testergebnissen in Erstsprache sogar übersteigen. Es kann somit festgehalten werden, dass höhere Maturitätsquoten mehrheitlich mit schlechteren Ergebnissen der in EVAMAR II erfassten Teile der Maturitätsausbildung verknüpft sind und sich damit die Frage der Chancengleichheit stellt.

Vergleiche nach Strukturelement Lang- oder Kurzzeitgymnasium

Ein Teil der Schweizer Gymnasien schliesst direkt an die Primarstufe an und integriert die Sekundarstufe I in ein Langzeitgymnasium. Es drängte sich deshalb auf zu untersuchen, ob sich Unterschiede in den Resultaten zu den Kurzzeitgymnasien ergeben. Der Vergleich zeigt, dass in fast allen Testbereichen und auch bei den Maturanoten die Maturandinnen und Maturanden, welche ein Langzeitgymnasium besuchten, signifikant besser abschneiden. Allerdings sind die Effektstärken für die Erstprache (ausser dem Teilbereich Grammatik) sowie für Mathematik gering. Bei den Biologietests (nicht aber bei den Biologienoten) bewegt sich die Effektstärke des Leistungsvorsprungs bereits in einem mittleren Bereich. Das Gesamtbild liesse sich damit erklären, dass wahrscheinlich vor allem Fächer mit einem systematischen Aufbau wie die Biologie (aber auch Chemie, Physik, Geschichte usw.) von der durchgehenden gymnasialen Unterrichtszeit bei einer Lehrperson mit einer fachspezifischen universitären Ausbildung sowie der strikten Fächerstruktur bereits auf der Sekundarstufe I profitieren und Lerneffizienzverluste infolge grosser Heterogenität der fachlichen Lernvoraussetzungen beim Wechsel von der Sekundarschule ins Kurzzeitgymnasium ausbleiben.

5 Ergebnisse III: Analyse der Maturaprüfungen

Im Projektteil D1 von EVAMAR II wurden die schriftlichen Maturaprüfungen der gleichen Stichprobe wie in Teilprojekt C (inklusive Tessin) qualitativ untersucht im Hinblick darauf, welche fachinhaltlichen und welche kognitiven Anforderungen sie stellen, und inwieweit es eine Übereinstimmung gibt mit den in den Teilprojekten A und B ermittelten Komponenten der Studierfähigkeit sowie mit den in Teilprojekt C durchgeführten Tests. Diese Analyse erfolgte zunächst für die Testfächer Erstsprache, Mathematik und Biologie; sie könnte in einem Folgeprojekt auf weitere Fächer ausgedehnt werden.

Die schriftlichen Schweizer Maturaprüfungen in Erstsprache zeigen sich in ausserordentlicher Breite und Vielfalt: Es scheint, als ob jede Schule und jeder Kanton, ja jede Lehrperson eigene Vorstellungen darüber hat, wie eine schriftliche Maturaprüfung auszugestalten sei und unter welchen Bedingungen die sogenannte „Hochschulreife“ überprüft werden soll. Die grösste Übereinstimmung hat die Schweizer Lehrerschaft darin, dass sie hauptsächlich Texte in der jeweiligen Originalsprache und selten Übersetzungen vorlegt, sowie darin, dass sie selten Texte mit weiblicher Autorschaft berücksichtigt. Ziemlich einig sind sich die Schweizer Lehrpersonen auch darin, dass sie meistens einen bis zwei Primärtexte aus dem 20. oder 21. Jahrhundert zu einem möglichst aktuellen oder zeitlosen Thema vorlegen. Dabei wählen sie zu rund 40-60% literarische Primärtexte aus und nur zu 10-20% Sachtexte; den Rest bilden sogenannte „Philosophische Texte“ (Deutsch 12%, Französisch 14%, Italienisch 30%). Vom Anspruchsniveau her gesehen, befinden sich die meisten der schriftlichen Maturaprüfungen (D 77%, F 99.1%, I 77%) im oberen Bereich der kognitiven Anforderungen (Niveau 4-6 nach Anderson u. a.). Die grössten regionalen Unterschiede liegen in der eindeutigen Dominanz des Anforderungsbereiches „Erschaffen-Bewerten“ in der Romandie (zwei Drittel), im Unterschied zum dritten Rang dieses Bereiches im Tessin (ein Fünftel). In der Deutschschweiz ist die Verteilung ausgeglichener, hier werden zu einem Drittel Beurteilungsleistungen gefordert, zu einem Viertel Analysen und zu einem weiteren Viertel das Erfinden von Produkten. Insgesamt stellt die vierstündige Aufsatzform ein Produkt der höchsten kognitiven Stufe, jener des Erschaffens, dar. Inhaltlich zeugen die vorgelegten Themen von einer grossen Aktualität, der Orientierung an der Lebenswelt der Jugendlichen, einer starken Betonung der Meinungsvielfalt und von einer intensiven Auseinandersetzung sowohl mit wichtigen gesellschaftlichen Fragen als auch mit zeitlosen Sinnfragen des Menschseins. Im Vergleich mit den Ergebnissen aus der Dozierendenbefragung stehen allerdings die ermittelten vielfältigen Prüfungsinhalte (Autorinnen und Autoren, Gesellschaftsthemen, ethische Probleme u. a.) im Grossen und Ganzen nur beschränkt im Erwartungshorizont der befragten Dozierenden aller Fachrichtungen. Das Leseverständnis wird zudem vor allem an literarischen und nur wenig an fachwissenschaftlichen Texten geprüft.

40.8% der untersuchten Aufgaben in Schweizer Biologie-Maturaprüfungen enthalten „higher-order questions“ (Hauptstufen 3-6 bei Anderson u. a.). Dieser Wert macht sie vergleichbar mit den bekannten, gross angelegten US-amerikanischen Assessments wie beispielsweise dem Medical College Admission Test (MCAT) oder dem „Advanced Placement (AP) Biology course. Und das ist aus zweierlei Hinsicht nicht selbstverständlich:

a) Deutlich später als im amerikanischen Kontext wurden erst in den 80er Jahren im deutschsprachigen Raum Ansätze zur Förderung des produktiven Denkens bzw. so genannte Denktrainings entwickelt. Erfahrungsgemäss kann es Jahrzehnte dauern, bis sich selbst empirisch gut abgesicherte universitäre Überlegungen in der tatsächlichen Lehrerbildung niederschlagen. Gemäss den vorliegenden Ergebnissen scheint der Transfer in die schweizerische SII-Didaktikausbildung aber grossflächig abgeschlossen zu sein und in vielen gymnasialen Schulzimmern Früchte zu tragen.

b) Sowohl das Verfassen als auch die Korrektur von guten Transferaufgaben ist heikel und sehr zeitintensiv. Offenbar sind viele Lehrpersonen davon überzeugt, dass sich der Aufwand lohnt.

Der hohe Prozentwert widerlegt auch das altbekannte Klischee, wonach in Biologie durch blindes Auswendiglernen leicht eine gute Note zu erreichen und „Denken“ unnötig sei. Die Resultate der Analyse der Biologieprüfungen weisen aber auch auf eine möglicherweise ungenügende Berücksichtigung der Themenbereiche „Ökologie und Systematik/Evolution“ hin.

Auch bei den Mathematikprüfungen liegt inhaltlich eine grosse Vielfalt vor. Die Maturaaufgaben zur Mathematik sind in aller Regel sorgfältig konstruiert und sehr phantasievoll ausgearbeitet. Es gibt daneben aber auch eine stattliche Zahl von „Standardprüfungen“: je eine Aufgabe zur Kurvendiskussion, zur Vektor- und zur Wahrscheinlichkeitsrechnung, zur Differentialrechnung (Extremalproblem) sowie eine Aufgabe mit mehreren Kurzaufgaben, die weitere Gebiete (Folgen und Reihen, elementare Algebra etc.) abdeckt. Bezüglich des kognitiven Anspruchsniveaus ist bei diesen Prüfungen der Anteil an „Erinnern“ besonders hoch. Grosse Unterschiede gibt es bei den (potentiellen) kognitiven Anforderungen sowohl zwischen Grundlagenfachprüfungen und Schwerpunktfachprüfungen als auch zwischen Grundlagenfachprüfungen verschiedener Schulen. Hinsichtlich des kognitiven Anspruchspotenzials ist der Grossteil der Matura(teil-)aufgaben der Kategorie „Ausführen; eine Prozedur auf ein bereits vertrautes Problem anwenden“ zuzuordnen. Die fast in jeder Prüfung vorkommende Extremalaufgabe gehört unserer Einschätzung nach meist in die Kategorie „Ein neues Problem auf ein bereits bekanntes Problem zurückführen“. Bei anspruchsvollen Extremalproblemen ist zudem ein hohes Mass an Modellierungsarbeit zu leisten, was dem Niveau „Modellieren, bekannte mathematische Modelle einsetzen“ (falls das Modell bereits bekannt ist) oder aber dem Niveau „Modellieren, neue mathematische Modelle einsetzen“ entspricht. Solche Aufgaben kommen fast nur in Prüfungen im Schwerpunktfach „Physik und Anwendungen der Mathematik“ vor.

6 Ergebnisse IV: Analyse der Maturaarbeiten

Dieses Teilprojekt wurde vollständig durch die Pädagogische Hochschule der Fachhochschule Nordwestschweiz konzipiert und durchgeführt.

Die Analyse der Qualität der Maturaarbeiten aus der Deutschschweiz, der Romandie und dem Tessin basierte auf drei Elementen: Im Zentrum der Untersuchung stand die Analyse von 437 Maturaarbeiten, die jeweils von zwei unabhängigen Expertinnen oder Experten einem Rating unterzogen wurden. Die Verfasserinnen und Verfasser dieser Arbeiten waren zudem Teil einer Stichprobe von Maturandinnen und Maturanden, die schriftlich zum Kontext, zum Verlauf und zu den Resultaten ihrer Projekte befragt wurde. Ergänzt werden diese Befunde durch die Rekonstruktion institutioneller Rahmenbedingungen auf nationaler und kantonaler Ebene sowie auf der Ebene von 30 an der Untersuchung beteiligten Gymnasien.

Die Evaluation der Maturaarbeiten kommt insgesamt zu einem positiven Befund: Es zeigt sich, dass die grosse Mehrheit der Maturandinnen und Maturanden mit den Grundregeln wissenschaftlichen Arbeitens vertraut ist. Die Analyse erhärtet die Hypothese, dass die inhaltliche Qualität der Arbeiten mit ihrer formalen und ihrer sprachlichen Qualität korreliert: Schülerinnen und Schüler, die bei den inhaltlichen Gütekriterien gut abschneiden, erzielen in der Regel auch bei den anderen Gütekriterien gute Werte, hingegen sind bei inhaltlich schwachen Arbeiten auch überdurchschnittlich viele formale und sprachliche Mängel festzustellen. Es wirkt sich positiv auf die Qualität der Maturaarbeiten aus, wenn die Schulen die Themenwahl durch die Vorgabe von Rahmenthemen strukturieren; fehlt eine solche institutionelle Beschränkung, so fallen die Arbeiten im Durchschnitt etwas schlechter aus. Dies dürfte nicht zuletzt darauf zurückzuführen sein, dass mit der Themenbreite auch die Komplexität der Betreuungsaufgaben zunimmt.

Zwischen natur-, sozial- und geisteswissenschaftlich ausgerichteten Arbeiten sind keine gravierenden qualitativen Differenzen festzustellen. Auch beim Geschlechtervergleich ergeben sich keine bedeutsamen Qualitätsunterschiede. Anders verhält es sich beim Regionalvergleich: Wiewohl der Befund insgesamt für alle Landesteile positiv ausfällt, stechen doch beträchtliche sprachregionale Unterschiede zwischen der Deutschschweiz und der lateinischen Schweiz ins Auge, die statistisch im Rahmen dieses Projekts nicht vollends aufgeklärt werden konnten. Es ist zu vermuten, dass das bessere Abschneiden der Romandie und des Kantons Tessin durch engere Vorgaben bei der Themenwahl und durch die Einbettung der Arbeit in seminarähnliche Strukturen begünstigt wird. Es kann jedoch nicht ausgeschlossen werden, dass die Ergebnisse des Ratings durch unterschiedliche sprachregionale Bewertungstraditionen beeinflusst worden sind.

In Hinblick auf die Rahmenbedingungen ist festzustellen, dass die Maturandinnen und Maturanden schulische Leitfäden und andere offizielle Papiere eher beiläufig zur Kenntnis nehmen; Transparenz und Verhaltenssicherheit ergibt sich für sie am ehesten auf der Basis interpersonalen Kontakts mit den betreuenden Lehrpersonen. Dabei ist weniger die Art und Frequenz der Kontakte entscheidend, sondern vielmehr die wahrgenommene Intensität der Unterstützung bei der Bewältigung kritischer Phasen. Dies beginnt bei der Formulierung des Themas und erstreckt sich über die Klärung von Methodenfragen bis hin zum formalen Aufbau des Berichts. Ähnlich wie die Frequenz der Kontakte mit Lehrpersonen ist auch die Bearbeitungsdauer der Projekte für die Qualität der Maturaarbeiten nur wenig bedeutsam.

In der Analyse fällt auf, dass das Rating der Expertinnen und Experten in der Regel kritischer ausfällt als die Benotung durch die Lehrpersonen. Dabei ist zu berücksichtigen, dass der Fokus der pädagogischen Bewertung sich vom Fokus der vorliegenden Untersuchung in mehrfacher Hinsicht unterscheidet. So befinden die Lehrpersonen in der Regel nicht allein über das schriftliche Produkt, sondern auch über die mündliche Präsentation. Zudem musste sich die vorliegende Qualitätsanalyse auf eine summative Dimension beschränken, da sie nicht prozessbegleitend angelegt werden konnte, während die Lehrpersonen auch den Lernfortschritt der Schülerinnen und Schüler im Projektverlauf vor Augen haben. Darüber hinaus ist zu betonen, dass die im Rahmen dieser Untersuchung durchgeführte Evaluation der Maturaarbeiten sich weitgehend auf den akademischen Blickwinkel des Erwerbs der Studierfähigkeit beschränkt, während die Lehrpersonen bei der Wahrnehmung ihres Bildungsauftrags eine Vielzahl anderer Aspekte in die Bewertung mit einfliessen lassen können.

Die Schülerinnen und Schüler tendieren dahin, der Maturaarbeit sowohl einen intrinsischen wie auch einen extrinsischen Wert zuzusprechen; nur in vergleichsweise wenigen Fällen wird das Projekt im Rückblick als nutzlos erachtet. Die inhaltliche Qualität korreliert positiv mit den Einschätzungen der Schülerinnen und Schüler zum Nutzen der Arbeiten für die weitere akademische oder berufliche Qualifikation wie auch mit dem Nutzen für die persönliche Entwicklung.

Von der Warte tertiärer Bildungsinstitutionen aus ergibt sich für die Maturaarbeit prima facie das Bild, dass die Maturandinnen und Maturanden in der Lage sein sollten, den elementaren Anforderungen eines Hochschulstudiums gerecht werden zu können: Sowohl bei der Strukturierung von Texten wie auch bei der Verwendung von Zitaten und beim Anlegen von Bibliographien legen die Maturaarbeiten davon Zeugnis ab, dass die gängigen akademischen Regeln bereits im Gymnasium eingeübt werden. Dasselbe gilt für die Wiedergabe elementarer

theoretischer und empirischer Sachverhalte sowie ansatzweise auch für die Gewinnung von Informationen durch Recherchen oder durch Experimente, wiewohl bei den schulischen Projekten deskriptive und reproduktive Anteile deutlich gewichtiger sind als analytische Anteile.

Indes ist hier zu betonen, dass eine Querschnittsanalyse zu den Maturaarbeiten keine Prognose im Hinblick auf die Nachhaltigkeit der Lernprozesse erlaubt. Die Reichweite der institutionellen Vorkehrungen, die Intensität der pädagogischen Betreuung und die Unterstützung durch das soziale Umfeld bilden Kontextfaktoren, deren Zusammenwirken eine hohe Qualität des Outcome begünstigt. Wirken diese Faktoren im Rahmen des Studienalltags nicht mehr gebündelt zusammen, ist nicht auszuschliessen, dass viele der ersten Studienarbeiten von Anfangssemestern hinter dem Niveau der Maturaarbeiten zurückbleiben.

7 Grenzen der Studie EVAMAR II

An dieser Stelle soll nochmals ausdrücklich auf die schon im zweiten Abschnitt erwähnten Grenzen der Studie EVAMAR II hingewiesen werden. Grundsätzlich können nur über einige ausgewählte Aspekte der Ziele und der Zielerreichung der gymnasialen Bildung Aussagen gemacht werden. Insbesondere folgende Komponenten einer umfassenden Studierfähigkeit wurden im Projekt EVAMAR II nicht untersucht und ihr Vorhandensein bei Maturandinnen und Maturanden nicht analysiert: erstens Wissen und Können, welches zwar Teil einer breiten Allgemeinbildung ist – als Grundlage zur Lösung anspruchsvoller Aufgaben in der Gesellschaft (vertiefte Gesellschaftsreife) – und damit in der gymnasialen Bildung vermittelt werden muss, aber an der Universität nicht direkt vorausgesetzt wird; zweitens zur Studierfähigkeit gehörendes notwendiges Wissen und Können aus anderen Unterrichtsfächern als Erstsprache, Mathematik und Biologie; drittens Elemente überfachlicher Kompetenzen, die im gegebenen Projektrahmen nur schlecht erfassbar gewesen wären, dennoch aber für die erfolgreiche Bewältigung eines Studiums wichtig sind (z. B. die Fähigkeit des selbstorganisierten Lernens, der allgemeinen Selbstorganisation, des Recherchierens, der raschen Verarbeitung umfangreicher Fachliteratur usw.). Die Bedeutung solcher Faktoren liesse sich nur durch weitere, im umfassenden Konzept vorgeschlagene Untersuchungen ermitteln (Längsschnittuntersuchungen bei Studierenden), und sie wären nur durch aufwändige Assessmentverfahren zu messen. Immerhin wurden solche Faktoren aber bereits im Projekt EVAMAR I über Befragungen, also über Selbsteinschätzungen und damit annäherungsweise valide ermittelt. Im Zusatzfragebogen von EVAMAR II wurden zudem diesbezüglich ergänzende Fragen zu EVAMAR I eingestreut.

VIII Schlussfolgerungen und Empfehlungen des Projektleiters

In diesem Kapitel werden wichtige Schlussfolgerungen gezogen und Empfehlungen aus der Sicht des Projektleiters formuliert, die auch normative Wertungen enthalten.

1 Zufriedenstellendes Gesamtergebnis

Die Schweizer Maturandinnen und Maturanden erzielten in den im Projekt EVAMAR II untersuchten Bereichen Ergebnisse, die für standardisierte, für eine bestimmte Population entwickelte Tests normal sind. Zu dieser Normalität gehört, dass die Aufgaben im Mittel etwa zur Hälfte richtig gelöst wurden. Die Fähigkeiten der Maturandinnen und Maturanden liegen im Durchschnitt auf einem Niveau, das die Lösung von Aufgaben eines – an universitären Anforderungen orientierten – mittleren Schwierigkeitsgrades erlaubt. Insofern kann der Ausbildungsstand der Schweizer Maturandinnen und Maturanden als zufriedenstellend bewertet werden.

Die Möglichkeiten und Grenzen der Leistungsmessung, insbesondere auch des gewählten testtheoretischen Modells, schliessen die Konstruktion eines Tests aus, bei dem *alle* Aufgaben richtig gelöst werden müssen, um die gemessene Kompetenz, hier Aspekte der Studierfähigkeit, nachweisen zu können. Abgesehen von der kaum durchführbaren Auswertung eines solchen Tests mittels der in der wissenschaftlichen Testtheorie bekannten Verfahren, würden entsprechende Aufgaben wohl von aussen als zu leicht bewertet und nicht ernst genommen. Aber auch bei „normal“ ausgefallenen Ergebnissen lässt sich nicht präzise bestimmen, ab welchem Punktwert Studierfähigkeit im entsprechenden Bereich gegeben ist. Bestehens- und Zulassungslimiten, wie sie z. B. bei den Bestehensnormen für Maturitätsprüfungen oder bei Zulassungsprüfungen gesetzt werden, können dies ebenso wenig leisten. Es gibt kein deterministisch genaues Messverfahren psychologischer Merkmale. Studierfähigkeit ist zudem keine absolute, stabile Eigenschaft, die sich dichotomisieren lässt. Ab welcher Grenze bei den in EVAMAR II eingesetzten Leistungstests ein erfolgreiches Studium nicht mehr möglich ist, kann also nicht deterministisch bestimmt werden.

Wahrscheinlichkeits-Prognosen über den Studienerfolg von Personen auf der Grundlage von psychometrischen Messinstrumenten lassen sich erst treffen, wenn die Prognosevalidität von Messinstrumenten mittels Vergleich mit dem effektiven Studienerfolg ermittelt worden ist. Solche Untersuchungen dauern nur schon deshalb mehrere Jahre, weil die Messzeitpunkte so weit auseinander liegen. Aber auch dann sind keine deterministischen Voraussagen möglich; bisherige Untersuchungen haben Korrelationen bis maximal $r = .5$ aufgezeigt (siehe auch die entsprechenden Beschreibungen in den Kapiteln I und IV). Für die in EVAMAR II eingesetzten Messverfahren konnten innerhalb der verfügbaren Zeit noch keine Prognosevalidierungen vorgenommen werden. Aufgrund ihrer theoretischen Fundierung und der konsequenten Ausrichtung der Testaufgabenentwicklung auf die Schnittstelle Gymnasium – Universität ermöglichen sie aber begründete Interpretationen der Testergebnisse über die oben festgestellte „Normalität“ hinaus. Weitergehende Aussagen, wie z. B., mit wie hoher Punktzahl in den verschiedenen Tests mit welcher Wahrscheinlichkeit mit einem bestimmten Studienerfolg zu rechnen ist, könnten erst nach weiteren, umfassenden und langwierigen empirischen Untersuchungen getroffen werden.

Die gefundene breite Streuung der Testresultate bedeutet gleichzeitig auch, dass die Gymnasien nicht alle ihre Maturandinnen und Maturanden mit Kompetenzen entlassen, die in der ganzen Breite als mindestens genügend eingeschätzt werden können. Das ist die Bestätigung eines Befundes, der aufgrund der effektiv an den Gymnasien vergebenen ungenügenden Noten an Maturitätsprüfungen vor allem in den Bereichen Mathematik und schriftliche Erstsprachprüfung schon bisher vermutet werden konnte. Auch die mehr oder weniger hohen Durchfallquoten in universitären Prüfungen sind unter anderem damit erklärbar, dass zumindest ein Teil der nicht erfolgreichen Studierenden nicht „allgemein studierfähig“ war. Das bedeutet aber auch, dass trotz befriedigendem Ausbildungsstand der Schweizer Maturandinnen und Maturanden Verbesserungen grundsätzlich möglich sind. Das ist keine aussergewöhnliche Schlussfolgerung, lässt sich eine solche doch für jede Schulstufe nach

einer Leistungsmessung mit standardisierten Tests ziehen. Zudem gelingt es wohl keiner selektiven Schulstufe, ihre Absolventen so zu entlassen, dass alle in der nächsten Schulstufe reüssieren.

Zwingender Handlungsbedarf würde dann bestehen, wenn die Aufgaben zu klar weniger als der Hälfte richtig gelöst und der Durchschnitt der geschätzten Personenfähigkeiten klar unter einem mittleren Wert liegen würden. Das ist bei den Schweizer Maturandinnen und Maturanden nicht der Fall.

Die Frage, was man nach Abschluss eines Bildungsganges behält und wie viel man vergisst, ist nicht gut untersucht. Im Biologietest haben die Maturandinnen und Maturanden, die dieses Fach bereits früher abgeschlossen hatten, zwar nicht alles vergessen, aber klar schlechter abgeschnitten. Zum Zeitpunkt des Abschlusses in Biologie war ihr Wissen und Können vermutlich noch umfassender als zum Testzeitpunkt, der unmittelbar vor der maturitären Qualifikation lag. Wenn an der Universität zuweilen lückenhaftes Eingangswissen beklagt wird, muss deshalb die Ursache nicht zwingend in einer (zu) tiefen Stundendotation des gymnasialen Faches liegen, sondern kann sich auch durch einen frühen Abschluss begründen. Es ist deshalb zu empfehlen, alle Fächer möglichst spät abzuschliessen, auch wenn eine breite Fächerung bis zur Matura andere Nachteile mit sich bringt.

2 Massnahmen zur Sicherung der allgemeinen Studierfähigkeit

Interpretiert man allgemeine Studierfähigkeit dahin gehend, dass jeder Maturitätsausweis genügende Eingangskompetenzen für alle möglichen Studienfächer bescheinigen sollte, so ist dies vermutlich für einen beachtlichen Teil der Maturandinnen und Maturanden nicht gegeben. Im Projekt EVAMAR II konnte das zwar nur für die getesteten Bereiche festgestellt werden, es gibt aber keinen Grund anzunehmen, dass sich das Bild eines ansehnlichen Anteils ungenügender Leistungen bei Einbezug weiterer Fachbereiche verändern würde. Welchem Prozentwert dieser Anteil entspricht, lässt sich nicht genau bestimmen (siehe die Überlegungen in Abschnitt 1).

Auch dieser Befund ist nicht neu, drücken sich doch diese Lücken – wie schon erwähnt – auch in den Maturanoten in Erstsprache und vor allem in Mathematik aus. Wenn die Leistungen der Maturandinnen und Maturanden an den schriftlichen Maturaprüfungen zu einem Fünftel (Erstsprache) oder gar mehr als zwei Fünfteln (Mathematik) als ungenügend bewertet werden, sind erhebliche Kompetenzmängel in diesen Fächern evident. Dass trotzdem allen Maturandinnen und Maturanden durch die Ausstellung des Maturitätszeugnisses allgemeine Studierfähigkeit attestiert wird, ermöglicht das Notenkompensationssystem des MAR. Damit sind auch die bekannten (subjektiven) und durch die Befragung in Teilprojekt A4 erhärteten Klagen von Universitätsdozierenden über Lücken vor allem in Erstsprache und Mathematik verständlich sowie die in einigen Studienfächern erheblichen Durchfallquoten in den ersten Prüfungen teilweise erklärbar. Ungeklärt bleibt allerdings, ob es sich in den Maturitätsprüfungen bei den für die ungenügenden Ergebnisse ursächlichen Teilen um Wissen und Können handelt, das für die allgemeine Studierfähigkeit unabdingbar ist. Im Extremfall könnten Maturandinnen und Maturanden mit ungenügenden Noten trotzdem in für die Studierfähigkeit relevanten Bereichen genügende Werte aufweisen, wie es umgekehrt sein könnte, dass Maturandinnen und Maturanden mit genügenden Noten nicht über genügende Kompetenzen in für die allgemeine Studierfähigkeit relevanten Bereichen des Fachs verfügen. Die höchstens mittelmässigen Korrelationen der Testergebnisse mit den Maturanoten zeigen jedenfalls, dass diese nur teilweise dasselbe messen wie die in EVAMAR II eingesetzten Tests (siehe auch die Vergleiche in Kapitel V). Aber auch die Ergebnisse der Tests, die durchwegs Wissens- und Könnens-Teile untersuchen, die für mindestens ein Studienfach relevant sind, fielen teilweise schlecht aus. Die Vermutung, dass nicht alle Maturandinnen und Maturanden in der ganzen Breite über genügende Kompetenzen verfügen, um jedes beliebige Studium aufnehmen zu können, ohne zusätzlich Lücken füllen zu müssen, lässt sich deshalb – unter Einbezug der Maturanoten – auf gute Gründe stützen. Dies, obwohl sie durch die Verleihung des Matura-Ausweises als „allgemein studierfähig“ qualifiziert wurden.

Die Grundfrage lautet deshalb, ob sich künftig eine „lückenlose Studierfähigkeit“ aller Maturandinnen und Maturanden erzielen bzw. sicherstellen lässt. Abgesehen vom Fehlen eines „treffsicheren“ Messinstrumentariums sind verschiedene Massnahmen denkbar, die im Folgenden genannt und diskutiert werden.

- Variante 1: Zunächst könnte man die Bestehensnormen an den Maturaprüfungen insofern revidieren, als ungenügende Noten nicht mehr zugelassen würden.

  Das hätte aber vermutlich zwei unerwünschte Effekte: Weil sich die kognitiven Leistungen von Schülerinnen und Schülern nicht beliebig verbessern lassen und am Gymnasium vermutlich bereits die diesbezüglich Besten ausgewählt worden sind, würde bei Beibehaltung der jetzigen Bewertungsmassstäbe die Maturitätsquote sinken. Es ist nicht zu erwarten, dass die jetzt 25% Ungenügenden in Mathematik sich auf genügende Noten verbessern könnten. Um einen solchen drastischen Rückgang zu vermeiden, würde wohl die Notengebung durch eine Lockerung der Bewertungsmassstäbe angepasst. Es bleibt ferner offen, ob die Notenqualifikation in den verschiedenen Fächern auch gleichbedeutend ist mit der Qualifikation der für die Studierfähigkeit notwendigen Kompetenzen aus diesen Fächern (siehe die Überlegungen dazu im 2. Absatz dieses Abschnitts). Dabei sei daran erinnert und sollte im Auge behalten werden, dass Studierfähigkeit nicht das einzige Ziel der gymnasialen Bildung ist.

- Variante 2: Man könnte mittels eines methodisch erweiterten Ansatzes von EVAMAR II bestimmen, welche Kompetenzen aus den verschiedenen gymnasialen Fächern für die allgemeine Studierfähigkeit wirklich wichtig sind und für diese die Nichtkompensierbarkeit ungenügender Leistungen festsetzen. Diese Kompetenzen müssten bei den Maturaprüfungen mit einer eigenen Bewertung versehen werden.

  Abgesehen vom enormen Aufwand für diese Arbeiten – von der Anforderungsanalyse bis zur Testentwicklung –, den damit verknüpften methodischen Schwierigkeiten und allenfalls unerwünschten Signalwirkungen für das Lernen („Abwertung“ von im Hinblick auf die Studierfähigkeit weniger relevanten Fachbereichen), würde ein derartiges Bestehensreglement wahrscheinlich ebenfalls zu einer politisch unerwünschten Reduktion der Maturitätsquote führen.

  Das Ziel einer „lückenlosen Studierfähigkeit“ für alle Maturandinnen und Maturanden bei gleichbleibender Maturitätsquote dürfte wohl eine Fiktion sein. Sie ist aber in unserer Einschätzung auch keine absolute Notwendigkeit, weil viele Maturandinnen und Maturanden durchaus in der Lage sein dürften, sich in Teilbereichen und in einem beschränkten Umfang das zum Zeitpunkt der Ausstellung des Maturazeugnisses noch fehlende Wissen und Können – bei Bedarf bzw. bei entsprechender Studienwahl – auch später anzueignen. Voraussetzung dazu ist selbstredend die Fähigkeit des selbstständigen Lernens.

- Variante 3: Man könnte die Aufgabe des Anspruchs der allgemeinen Studierfähigkeit mit gleichzeitiger Aufgabe der allgemeinen Zutrittsberechtigung für Inhaberinnen und Inhaber einer Schweizer Matura zu den Universitäten erwägen. Das wäre gleichbedeutend mit einer Begrenzung auf eine „Fakultätsreife“, für die sowohl bezüglich ihres Erwerbs am Gymnasium wie auch bezüglich ihrer tatsächlichen Berechtigungs-Ausstattung verschiedene konkrete Varianten vorstellbar sind. Am Gymnasium wäre z. B. ein zweistufiges System denkbar, in dem die Spezialisierung erst in einer zweiten Stufe im letzten oder in den beiden letzten Ausbildungsjahren erfolgt. Der Maturitätsausweis würde dann nur für den Zutritt zu einer bestimmten Fakultät berechtigen. Noch weiter ginge die Vorgabe einer auch formalen Spezialisierung in genau dem einen späteren Studienfach, wie das beispielsweise in gewissen angelsächsischen Systemen der Fall ist. Es wäre klar, dass dann kein Fach bei ungenügenden spezifischen Voraussetzungen studiert werden dürfte.

  Auch bei einer solchen Lösung wäre die Festlegung der „fakultätsspezifischen“ Anforderungen nicht problemlos. Die Untersuchungen in den Teilprojekten A haben gezeigt, dass es keine lineare Aufbau-Verknüpfung zwischen Maturitätsfächern und Studienfächern gibt, sondern für jedes Studienfach Wissen und Können aus einem mehr oder weniger breiten Fächer-Fundament der Gymnasien erforderlich ist. Der Schritt zur Verlagerung der Prüfungs- bzw. Berechtigungskompetenz an die Universität läge dann nahe. Eine „Fakultätsreife“ würde zudem andere Ziele der gymnasialen Bildung, namentlich die „Vorbereitung auf anspruchsvolle Aufgaben in der Gesellschaft“, beschneiden (siehe nächster Abschnitt). Der Anspruch, mit dem Maturitätsausweis die allgemeine Hochschulreife mit Anrecht auf prüfungsfreien Zugang zu den Hochschulen zu bescheinigen, sollte deshalb nicht aufgegeben werden.

Ein wichtiges Argument für die Beibehaltung der allgemeinen Zugangsberechtigung trotz bereichsweise mangelhafter Fähigkeiten sind auch die geltenden Einschränkungen hinsichtlich der erlaubten ungenügenden Kompetenzbereiche. Die Maturandinnen und Maturanden mit ungenügenden Leistungen müssen sich immerhin über doppelte Besserleistungen in anderen Fächern ausweisen, und es sind höchstens drei Noten unter 4 möglich. Dies hat laut den Ergebnissen von EVAMAR I im Vergleich mit der Regelung vor der Maturitätsreform dazu geführt, dass die Misserfolgsquote sich zumindest für die Deutschschweiz beinahe verdoppelt hat (Ramseier et al., 2005, S. 139).

Ferner sind die praktischen Auswirkungen der festgestellten Kompetenzlücken zu berücksichtigen: Das Projekt EVAMAR I und weitere Untersuchungen haben aufgezeigt, dass eine Kanalisierung der Studienwahl aufgrund der eigenen Möglichkeiten und der damit verknüpften Interessen von selbst erfolgt und sich daher Lücken in gewissen Fächern nur beschränkt auf den konkreten Studienerfolg auswirken.

- Variante 4: Man sollte prüfen, ob es im Hinblick auf eine Revision des MAR sinnvoll ist, wenigstens für ausgewählte basale Kompetenzen bzw. Grund-Studierkompetenzen Bestehensnormen ohne die Möglichkeit der Kompensation ungenügender Ergebnisse für alle Maturandinnen und Maturanden festzulegen. Auf der Grundlage der Ergebnisse von EVAMAR II gehören Erstsprach- und Mathematikkompetenzen dazu. Die Grund-Studierkompetenzen machen nicht die ganze Breite der jeweiligen gymnasialen Fach-Curricula aus, sondern nur die für eine Vielzahl von Studienfächern Relevanten. Sie sind deshalb nicht gleichzusetzen mit den Kompetenzen, die in Erstsprache und Mathematik an den Maturaprüfungen gemessen werden, sondern umfassen nur einen Teil davon. Die weiteren, unter gymnasialen Zielsetzungen ebenfalls geprüften und durchaus ebenso wichtigen curricularen Inhalte würden bei ungenügenden Ergebnissen kompensierbar bleiben. Es handelt sich also nicht um eine allgemeine Forderung nach der einen Mindest-Maturanote 4 in Mathematik und Erstsprache. Wie bei Variante 2 müsste die Prüfung dieser Bereiche bei den Maturaprüfungen mit einer eigenen Bewertung versehen werden. Aus den Analysen von EVAMAR II lässt sich allerdings noch zu wenig genau bestimmen, welches die genauen Grund-Studierkompetenzen sind. Sie müssten nach den Regeln der Kompetenzmodell-Entwicklung noch präziser festgelegt werden. Dabei wäre – in Ergänzung zum Verfahren bei HarmoS – der analytische Ansatz zur Ermittlung vorausgesetzten Wissens und Könnens in Erstsprache und Mathematik, wie er in EVAMAR II praktiziert wurde, einzubeziehen, ebenso die Diskussion und verbindliche Einigung auf die gewollten und gesollten Grund-Studierkompetenzen durch die massgebenden Akteure an der Schnittstelle Gymnasium – Universität.

  Der bei Variante 2 befürchtete Aufwand für ein solches Verfahren bliebe natürlich bestehen, müsste aber nicht mehr für die ganze Breite aller Maturitätsfächer betrieben werden. Die Nichtkompensierbarkeit von ungenügenden Kompetenzen in nur wenigen und nicht ganze Fächer umfassenden Bereichen wäre wohl auch im Hinblick auf ein mögliches Absinken der Maturitätsquote im Ausmass verkraftbar.

Der Projektleiter befürwortet die Massnahmenvariante 4. Sie würde dem Anspruch der maturitären Qualifikation für die allgemeine Studierfähigkeit wieder mehr Boden verschaffen. Man kann auch davon ausgehen, dass absolute Bestehensvorgaben ohne Kompensationsmöglichkeiten eine bessere Steuerwirkung insofern entfalten würden, als eine vorzeitige Resignation in diesen Bereichen durch die Maturandinnen und Maturanden weniger schnell „kalkuliert“ würde. Da Englisch als Wissenschaftssprache ebenfalls zu den Basiskompetenzen der Studierfähigkeit gehört – was auch die in Teilprojekt A4 befragten Dozierenden so sahen –, sollte in gleicher Weise auch das Erreichen einer (noch zu bestimmenden) Standardkompetenz in Englisch obligatorisch werden.

3 Beibehaltung der breiten Fächerung und der allgemeinen Zutrittsberechtigung an Hochschulen

In Abschnitt 2 wurde bereits die Beibehaltung der allgemeinen Zutrittsberechtigung für Inhaberinnen und Inhaber einer Schweizer Maturität empfohlen. Sie muss auch unter dem Aspekt der zweiten Zielsetzung des Gymnasiums, der „vertieften Gesellschaftsreife" betrachtet werden.

Der Projektleiter bewertet aus normativer Perspektive die aktuell gültige, doppelte Zielsetzung der generellen Studierfähigkeit und der breiten Allgemeinbildung als richtig. Und zwar in der umgekehrten Gewichtung. Die Begründung ist die folgende: In unserer Gesellschaft müssen in der Zukunft eine Reihe gewichtiger Probleme gelöst werden. Welche Probleme in Angriff genommen und wie sie gelöst werden, wird zumindest in unserer Kultur demokratisch bestimmt. Das Gymnasium besuchen mehrheitlich die kognitiv leistungsfähigeren Schülerinnen und Schüler, von denen viele später nicht nur Staatsbürgerinnen und Staatsbürger sein werden, sondern auch verantwortungsvolle Positionen einnehmen und gestaltend wirken (können). Gerade für sie ist deshalb die gymnasiale Zielsetzung des Erwerbs der Kompetenzen zur Lösung anspruchsvoller Aufgaben besonders adäquat. Anspruchsvolle Aufgaben und Probleme sind zwar häufig fachübergreifend, in ihren Teilaspekten aber letztlich immer fachspezifisch. Zudem herrscht heute – in Abgrenzung zu den veralteten formalen Bildungstheorien und deren neueren Varianten (wie etwa die Schlüsselqualifikationstheorien) – weitgehend wissenschaftlicher Konsens darüber, dass zur Lösung sowohl fachübergreifender wie auch fachspezifischer Probleme die Verfügbarkeit von entsprechendem fachspezifischem Wissen und Können von ausschlaggebender Bedeutung ist. Das Wissen über allgemeine Methoden z. B. ist dazu ebenso wenig ausreichend wie die Bildung in einem fachfremden Bereich. Deshalb sind Bildungsinhalte mit dem Ziel der vertieften Gesellschaftsreife im Hinblick auf potenzielle Problemstellungen nicht beliebig, sondern sie müssen sorgfältig und zielbezogen ausgewählt und selbstverständlich in einem aktiven Lernprozess, der Zeit erfordert, erarbeitet werden. Gesellschaftlich relevante Probleme stellen sich in vielen Fachbereichen; es braucht deshalb eine breit gefächerte Bildung. Eine Abkehr vom Ziel der allgemeinen Studierfähigkeit hin zur Ausrichtung auf eine Fakultätsreife konfligiert mit dem Ziel der vertieften Gesellschaftsreife durch eine breite Allgemeinbildung.

Das Offenhalten des allgemeinen Zugangs zu allen Studienfächern und die damit verbundene breite Fächerung des Gymnasiums hingegen harmonieren mit dem Ziel der vertieften Gesellschaftsreife. Das Problem der genauen Bestimmung der Fächer und deren angemessene Curricula angesichts einer beschränkt verfügbaren Lernzeit am Gymnasium ist damit allerdings noch nicht gelöst. Die Untersuchungen von EVAMAR II bestätigen bisherige Befunde (z. B. Huber, 1998 und im Druck), dass zu einer allgemeinen Studierfähigkeit auf jeden Fall Grund-Studierkompetenzen aus Erstsprache und Mathematik, aber auch – das stützt die Befragung der Dozierenden – aus der Wissenschaftssprache Englisch gehören. Diese Grund-Studierkompetenzen füllen aber nur einen Teil der gymnasialen Curricula und Prüfungen für diese Fächer aus, wie EVAMAR II besonders für Erstsprache zeigen konnte. Diese Curricula enthalten immer auch mehr oder weniger weite Bereiche, die durch das Ziel der vertieften Gesellschaftsreife begründet sind. Wie weit in den gymnasialen Curricula unter diesen Zielvorgaben (allgemeine Studierkompetenz und vertiefte Gesellschaftsreife) auch nicht notwendige Teile enthalten sind, muss offen bleiben; eine solche Untersuchung gehörte nicht zum Auftrag von EVAMAR II. Die Analysen in den Teilprojekten A1 sowie A3 und A4 haben im Weiteren ergeben, dass die meisten untersuchten Studienfächer auch spezifisches Wissen und Können aus ihrem Fach voraussetzen. Wenn aber alle Studiendisziplinen Spezialwissen der Studienanfängerinnen und -anfänger in ihrem Fach voraussetzen würden, wäre das Gymnasium schnell überfordert. Gemäss Art. 5 des MAR sollen die Maturitätsschulen einer solchen allfälligen Forderung auch nicht entsprechen: am Gymnasium soll ausdrücklich keine fachspezifische Ausbildung angestrebt werden. Auch in der Begründung zum Rahmenlehrplan heisst es, das Gymnasium könne und dürfe nicht „Propädeutik für einzelne Wissenschaftszweige betreiben" (EDK, 1994, S. 6). Es ist deshalb geradezu ein Gebot, die fachspezifischen

Eingangs-Anforderungen auf „grundlegende“ Kenntnisse einzugrenzen. Das gilt auch für jene Studienfächer, für deren gymnasiale „Tochterfächer“ in der aktuellen bildungspolitischen Diskussion mit dem Argument der Verbesserung der Anschlussfähigkeit an die Universität eine Stärkung gefordert wird. Welches diese grundlegenden Kenntnisse in welchen Studienfächern sind, ist eine immer noch nicht fundiert beantwortete Frage. Ein Verweis auf die bisherige Tradition ist aus unserer Sicht ein unzureichendes Argument; ebenso die naive didaktische Ableitung, wonach die grosse Bedeutung eines Studienfachs für die Gesellschaft verknüpft mit einer tiefen Studierendenquote einen Ausbau des gleichlautenden gymnasialen Faches erfordere. Es stellt sich deshalb gar die Frage, ob sich das Gymnasium nicht sogar *primär* am gebildeten Menschen ausrichten müsste, der anspruchsvolle Aufgaben in der Gesellschaft lösen kann. Die breite Fächerung wäre dann gewährleistet, und die Universitäten könnten auf jenen fachlichen Kompetenzen aufbauen, die unter dieser Zielsetzung zu erwerben sind. Damit würde auch der potenzielle Streit der Universitätsfächer um Stunden am Gymnasium entfallen.

4 Ausgleich der Defizite zu Beginn eines Studiums

Die empfohlene Beibehaltung der allgemeinen Zutrittsberechtigung ist in unserer Einschätzung nicht ohne fallweise fachliche Lücken bei zumindest einigen Maturandinnen und Maturanden praktizierbar. Das unvermeidliche Problem der nicht erzielbaren „lückenlosen Studierfähigkeit“ sollte deshalb durch ein grösseres Angebot von „Ausgleichskursen“ entschärft werden, die von den Universitäten studienfachbezogen angeboten werden. Zur Auffüllung von Lücken in Mathematik etwa existieren bereits Angebote an den Universitäten. Bei Umsetzung der Empfehlung in Abschnitt 2 würde allerdings die Notwendigkeit solcher Kurse für Mathematik und die anderen Fächer, für die nichtkompensierbare, basale Standards festgelegt wären, entfallen. Im Weiteren sollten sich die Verantwortlichen für die verschiedenen Studiengänge bei der Festlegung der fachlichen Studienvoraussetzungen trotzdem an die Vorgabe halten, dass eine spezifische Vorbereitung auf bestimmte Studienfächer nicht zum Auftrag des Gymnasiums gehört (siehe auch Abschnitt 3).

5 Massnahmen zur Förderung des selbstständigen Lernens ohne Abbau der Fachlichkeit

Die Ergebnisse der Dozierendenbefragung in Verbindung mit der Zusatzbefragung der Maturandinnen und Maturanden bestätigen bisherige Evaluationsuntersuchungen insofern, als die Fähigkeit zu selbstständigem Lernen und Arbeiten in der Selbsteinschätzung als teilweise lückenhaft beschrieben wird (Ramseier et al., 2005, Notter & Arnold, 2003 und 2006). Den Empfehlungen dieser Studien ist deshalb zuzustimmen: Der gezielten Förderung des selbstständigen Lernens an den Gymnasien sollte noch mehr Gewicht verliehen werden. Damit würde nicht nur eine der für das weitere Studium wesentlichen allgemeinen Kompetenzen gefördert, sondern Maturandinnen und Maturanden auch besser in die Lage versetzt, Eingangslücken, die je nach gewähltem Fachstudium individuell vorhanden sein können und – wie bereits festgestellt – unvermeidbar sind, selbstständig aufzufüllen. Die verstärkte Förderung des selbstständigen Lernens und Arbeitens darf aber weder auf Kosten von Fachinhalten geschehen noch zu Lasten der Breite der Fächerung am Gymnasium. Sonst könnten wieder eher fachliche Lücken im Hinblick auf die breite Fächerung der universitären Studien entstehen, und das weitere – in EVAMAR II nicht untersuchte – wichtige Ziel des MAR (vertiefte Gesellschaftsreife durch breite Allgemeinbildung) könnte tangiert werden. Das bedingt eine noch bessere Konzentration auf das Wesentliche in allen Fächern, z. B. auf der Grundlage einer besseren Analyse der Notwendigkeiten im Hinblick auf das Ziel der vertieften Gesellschaftsreife und durch eine bessere Absprache mit den Hochschulen über die Eingangserwartungen.

Eine Einführung eines Faches „selbstständig Lernen lernen“ wird damit nicht empfohlen. Es gibt genügend missglückte Versuche aus der Schulpraxis (mangelnder Transfer in die Fächer, nicht fachadäquate Methoden, zu wenig ausdauernde Steuerung der notwendigen Lernprozesse und damit keine Überwindung der Durststrecke bis zur erfolgreichen Internalisierung durch die Schülerinnen und Schüler). Damit die Integration des „selbstständig Lernen lernen“ gelingt, muss die Unterrichtsdidaktik in den Fächern darauf ausgerichtet werden. Das hat weitreichende Massnahmekonsequenzen bis hin zur entsprechenden Aus- und Weiterbildung von Gymnasiallehrkräften.

6 Massnahmen zur Verbesserung der Erstsprachkompetenzen

Die Behebung der bei einem Teil der Maturandinnen und Maturanden festgestellten Defizite bei den Erstsprachkompetenzen ist nicht nur eine Aufgabe des gymnasialen Fachs Erstsprache. Denn die Mängel beziehen sich auf die Erstsprachanwendung, die bereits in den meisten anderen Fächern des Gymnasiums praktiziert wird sowie auch dort und nicht erst im Studium zu den vorausgesetzten Kompetenzen gehört. Eine Verbesserung der Erstprachkompetenzen kann deshalb durch eine konsequente Förderung korrekter Erstsprachanwendung in allen Fächern angezielt werden. Auch dieser Aspekt sollte in der Aus- und Weiterbildung von Gymnasiallehrpersonen in allen Fächern expliziert werden.

7 Vermehrte Standardorientierung bei den Unterrichtsprogrammen

Die gefundenen grossen Unterschiede nicht nur zwischen den Testergebnissen der Einzelpersonen, sondern auch zwischen jenen ganzer Klassen, können nicht nur in unterschiedlichen Zielerreichungsgraden, sondern auch in unterschiedlichen faktischen Curricula liegen. Der Schweizerische Rahmenlehrplan (RLP) ist zu wenig bindend. Die einzelnen kantonalen oder schulischen Lehrpläne können beträchtliche Unterschiede aufweisen, und sie entfalten teilweise bezüglich Vorgabe von Qualitätsstandards wenig Relevanz. Die analysierten Maturaprüfungen zeigen wohl auch deshalb grosse Unterschiede beim Anforderungsniveau. Während die mehrheitlich anforderungsreichen Prüfungen und die mehrheitlich befriedigenden bis ausgezeichneten Testleistungen keine verbindlicheren Vorgaben erfordern, empfehlen sich solche im Sinne der Festlegung von verbindlichen Minimalstandards. Damit sollen die gefundenen unerwünschten Ausschläge nach unten vermieden und die Exzellenz nach oben nicht behindert werden. Verbindliche Minimalstandards ermöglichen auch weiterhin inhaltliche Heterogenität im oberen Bereich.

8 Vermehrte Standardorientierung bei der Leistungsbewertung

Die gefundenen Unterschiede zwischen den Klassen sind teils sehr gross. Die Noten zeigen grosso modo in die gleiche Richtung wie die Testergebnisse, es gibt aber auch „Anomalien“. Gymnasiallehrkräfte orientieren sich bei der Bewertung offenbar recht gut an übergreifenden Gütestandards, es müssen aber auch Anpassungen an das Leistungsvermögen von Klassen vermutet werden. Das schmälert die Aussagekraft von Maturanoten. Es ist deshalb zu überlegen, wie eine vermehrte Standardorientierung gesichert werden kann. Dabei ist von zentralen Maturaprüfungen eher abzuraten, weil diese ihrerseits mit vielen, in der Literatur gut diskutierten problematischen Aspekten behaftet sind. Zumindest die objektivierte Bewertung des Erreichens von Minimalstandards sollte ermöglicht und gefördert werden. Als Massnahmen empfehlen sich die folgenden:

- In der Aus- und Weiterbildung von Lehrpersonen sollte dem standardorientierten Unterrichten und Prüfen grösseres Gewicht beigemessen werden.
- Es sollten für zentrale Leistungsmessungen Zwischenlösungen untersucht werden, die weniger zentralistisch als eine Zentralmatura sind und trotzdem eine wirksame Orientierung an überschulischen Standards ermöglichen. Dazu gehört die Methode, anhand eines auf Kompetenzmodellen aufgebauten, abrufbaren Testsystems den einzelnen Lehrkräften nach einer bestimmten Zeitspanne Unterricht (von einem Quartal bis zu zwei Jahren) im entsprechenden Fach Rückmeldungen darüber geben zu können, wie stark sie in der Zielerreichung von den Standardvorgaben abweichen. Diese Form der Evaluation wäre freiwillig und die Rückmeldungen würden ausschliesslich an die Lehrkräfte erfolgen. Entsprechende Erfahrungen aus dem Volksschulbereich (Klassencockpit und „Module“ im Kanton St. Gallen) lassen eine gute selbstregulierende Wirkung erwarten.
- Auch einzelne Schulen sollten sich an den universitären Erfolgsquoten ihrer Absolventinnen und Absolventen messen können. Für die Ermittlung dieses Erfolgskriteriums drängt sich die Nutzung der Hochschulstatistik auf. An den meisten Universitäten und an der ETH werden die Informationssysteme zurzeit ausgebaut. Es sollte deshalb möglich sein, zu folgenden Daten zu gelangen, die auch dem nationalen Monitoring des Gymnasiums dienen:
  - Erfassung der Bildungsprofilwahl am Gymnasium (Schwerpunkt- und Ergänzungsfächer, Themenbereich der Maturaarbeit) und der Maturitätsnoten auf der Grundlage der eingereichten Maturitätszeugnisse, verbunden mit der Studienwahl und dem Erfolg im Studium (Noten in den Prüfungen und sonstigen Leistungsnachweisen).
  - Systematische Erhebung von Bildungskarrieren durch systematische Erfassung des Studienprozesses (Anzahl Wechsel der Studienrichtung, Studienabbrüche, Studiumsdauer, Abschlussarten (Bachelor, Master, Ph.D. usw.)), um diese mit den Daten der Abschlüsse in Zusammenhang zu setzen.
  - Erfassung der ausschlaggebenden Beweggründe bei der Wahl der Studienrichtung und des Studienortes über eine standardisierte schriftliche Eingangsbefragung (z. B. Interesse, gute Noten im Gymnasium, Karriere, Sprache an der Uni, Kosten usw.).

9 Beibehaltung der vierjährigen Ausbildungsdauer am Schulort Gymnasium

Die Resultate von EVAMAR II deuten darauf hin, dass die Gymnasien mit nur dreijähriger Mindest-Dauer am Schulort Gymnasium benachteiligt sind. Offenbar gelingt es den dreijährigen Gymnasien weniger gut als den mindestens vierjährigen, zusammen mit den progymnasialen Stufen der örtlichen Sekundarschulen innerhalb der gleichen Zeit die gleich guten fachlichen und überfachlichen Wissensstrukturen und Kompetenzen der Schülerinnen und Schüler aufzubauen. Andere Fachbereichskonzeptionen der zwei Bildungsstufen, unterschiedliche Eingangsvoraussetzungen beim Eintritt ins dreijährige Gymnasium sowie das Fehlen des in der Gymnasiallehrkraft personalisierten und damit vermutlich kohärenteren Unterrichtsaufbaus im Fach über die beiden Stufen mögen Ursachen dafür sein. Der entsprechende Passus im MAR sollte überdacht werden. Zumindest sollten keine weiteren Kantone auf die dreijährige Mindest-Dauer wechseln, auch wenn die Gymnasien die teuerste Kategorie von Schulen sind. Die Daten von EVAMAR II zeigen auch, dass dort, wo das Untergymnasium geführt wird und damit die Dauer des Gymnasiums noch länger ist, das fachliche Niveau bei allen erhobenen Leistungsdaten (inkl. Maturanoten) am Schluss im Ausmass zwar nur leicht, aber signifikant höher liegt.

10 Sicherung der allgemeinen Studierfähigkeit bei allen Schwerpunkten

Die gefundenen Leistungsunterschiede zwischen den verschiedenen Schwerpunktgruppen sind teilweise erheblich. Sie zeigen sich nicht nur bei den Tests von EVAMAR II, sondern in abgeschwächter Form auch bei den von den Schulen selbst vergebenen Maturanoten. Auch wenn die praktischen Auswirkungen wegen der unter Abschnitt 2 angesprochenen, „angepassten" Studienwahl der Maturandinnen und Maturanden nicht so gravierend sein dürften, finden vermutlich im Hinblick auf das Ziel der allgemeinen Studierfähigkeit auch problematische Kanalisierungen statt. Zudem erfolgt der Kompetenzaufbau für ein Studienfach keineswegs ausschliesslich im zugehörigen gymnasialen Schwerpunktfach. Für Studierende der Wirtschaftswissenschaften z. B. bleiben als Voraussetzung für ein erfolgreiches Studium solide Kompetenzen in Fächern ausserhalb des Schwerpunkts Wirtschaft und Recht, namentlich in Mathematik und in Erstsprache wichtig.

Auch hier zeigen nicht erst die Testergebnisse die Leistungsunterschiede, sie sind auch bei den Maturanoten ersichtlich. Wenn bei den neuen Schwerpunkten teilweise mehr als 50% der Maturandinnen und Maturanden in der schriftlichen Maturaprüfung in Mathematik mit einer ungenügenden Note bewertet wurden, ist ihre allgemeine Studierkompetenz zu hinterfragen. Die vergebenen Maturanoten wurden aber offenbar noch nie schweizweit systematisch ausgewertet.

Die Resultate von EVAMAR II können also die in EVAMAR I (Ramseier et al., 2005, S. 138) festgestellte und weiterzitierte. Gleichheit der Leistungen in den Grundlagenfächern für die Maturandinnen und Maturanden der – im Vergleich zu den alten Maturitätstypen – neuen Schwerpunkte nicht bestätigen. Die Ergebnisse sind teilweise schlechter ausgefallen. Die Studie EVAMAR II ermöglicht hier eine Korrektur der bisherigen Erkenntnisse. Es wäre nun aber falsch oder zumindest verfrüht, bessere und schlechtere Ergebnisse durch die Wirkung des Unterrichts in den entsprechenden Schwerpunktfächern ursächlich zu erklären. Unterschiede könnten bereits bei der Schwerpunktwahl bestehen, und die Dynamik des sich im Verlaufe des Gymnasiums entwickelnden Interaktionsgefüges von Lehren, Lernen und Bewerten kann sich klassenbezogen ganz unterschiedlich entwickeln. Es sollten aber Massnahmen zur Verbesserung der Ausbildungsergebnisse in studienrelevanten Teilbereichen der Fächer Mathematik und Erstsprache für einen Teil der Maturandinnen und Maturanden geprüft werden; dies betrifft vor allem jene, die einen mit dem MAR 95 – im Vergleich zu den alten Maturitätstypen – neu eingeführten Schwerpunkt gewählt haben. Die in Abschnitt 2 vorgeschlagene Aufhebung von Kompensationsmöglichkeiten ungenügender Leistungen in basalen Teilbereichen (Grund-Studierkompetenzen) gehört zu diesen Massnahmen. Eine Senkung der Bestehensquote um 25% oder mehr wäre wohl kaum zu befürchten. Viel eher würde eine solche Massnahme in unserer Einschätzung vor allem steuernde Wirkung auf die Lernanstrengungen im Verlaufe der Ausbildung am Gymnasium entfalten.

11 Annäherung der Maturitätsquoten

Die Hypothese, dass hohe Maturitätsquoten mit eher schlechteren Ergebnissen einhergehen, hat sich nicht nur bei den Tests, sondern grösstenteils auch bei den erhobenen Noten erhärtet. In den Kantonen mit hohen Maturitätsquoten sind die Aufnahmequoten an die Gymnasien ebenfalls höher. Weil es keinen Grund gibt anzunehmen, dass sich hier die Schulleistungen weniger gut entwickeln als in den Kantonen mit tiefen Maturitätsquoten, liegt die Vermutung nahe, dass die durchschnittliche Schulleistungsfähigkeit bereits beim Eintritt ins Gymnasium tiefer liegt und damit in den Kantonen mit hohen Maturitätsquoten auch eine höhere Zutrittschance zum Gymnasium bei gleicher Leistungsfähigkeit besteht. Zudem vermag die Ausbildung am Gymnasium diese Unterschiede offenbar nicht zu „heilen". Die offensichtlichen Chancenungleichheiten müssen politisch diskutiert werden.

12 Weitere Anstrengungen bei der Gender-Problematik

Die Unterschiede zwischen den Geschlechtern folgen teilweise dem bereits bekannten Bild. Die Bestrebungen zur Verbesserung der Leistungen von Gymnasiasten in Erstsprache und von Gymnasiastinnen in Mathematik und Naturwissenschaften bzw. der vermutlich dahinter stehenden Interessen sind weiterzuführen. Es sollten auch mögliche Unterschiede in weiteren Fächern, die nicht Gegenstand der Untersuchungen von EVAMAR II waren, untersucht werden.

13 Weiterentwicklung der Maturaarbeit

Die Einführung der Maturaarbeit hat sich bewährt. Die selbstständige Bearbeitung von Projekten als wichtiges Element gymnasialen Lernens sollte unter Einbezug der in den Untersuchungen von EVAMAR II gefundenen Erfolgsfaktoren in moderater Weise ausgebaut werden.

14 Weitere Untersuchungen

Das Projekt EVAMAR II ist als Querschnittsprojekt angelegt. Die Resultate werfen deshalb mehr Fragen nach den Ursachen der Befunde auf, als zuverlässig beantwortet werden können. Diesen offenen Fragen sollte mittels Längsschnittanalysen unter systematischer Ausnützung aller im Verlaufe der gymnasialen Ausbildung und des späteren Studiums anfallenden Personendaten nachgegangen werden.

Literaturverzeichnis (ohne Kapitel VI)

Abraham, U., Beisbart, O., Koß, G., & Marenbach, D. (2003). *Praxis des Deutschunterrichts. Arbeitsfelder, Tätigkeiten, Methoden*. Donauwörth: Auer.

Alexander, P. A., & Judy, J. E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research, 58*(4), 375-404.

Alexander, P. A., Pate, P. E., Kulikowich, J. M., Farrell, D. M., & Wright, N. L. (1989). Domain specific and strategic knowledge: Effects of training on students of differing ages or competence levels. *Learning and Individual Differences, 1* (3), 283-325.

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology, 80*, 260-267.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Raths, J., et al. (2001). *A taxonomy for learning, teaching and assessing. A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Artelt, C., & Schlagmüller, M. (2004). Der Umgang mit literarischen Texten als Teilkompetenz im Lesen? Dimensionsanalysen und Ländervergleiche. In U. Schiefele, C. Artelt, W. Schneider & P. Stanat (Hrsg.), *Struktur, Entwicklung und Förderung von Lesekompetenz. Vertiefende Analysen im Rahmen von PISA 2000* (S. 169-196). Wiesbaden: VS Verlag für Sozialwissenschaften.

Artelt, C., Stanat, P., Schneider, W., Schiefele, U., & Lehmann, R. (2004). Die PISA-Studie zur Lesekompetenz: Überblick und weiterführende Analysen. In U. Schiefele, C. Artelt, W. Schneider & P. Stanat (Hrsg.), *Struktur, Entwicklung und Förderung von Lesekompetenz. Vertiefende Analysen im Rahmen von PISA 2000* (S. 139-168). Wiesbaden: VS Verlag für Sozialwissenschaften.

Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. New York: Holt, Rinehart, and Winston.

Bamberger, R., & Vanecek, E. (1984). *Lesen–Verstehen–Lernen–Schreiben. Die Schwierigkeitsstufen von Texten in deutscher Sprache*. Wien: Jugend und Volk, Diesterweg, Sauerländer.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191–215.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.

Baumert, J., Bos, W., & Lehmann, R. (Hrsg.). (2000). *TIMSS/III: Dritte Internationale Mathematik- und Naturwissenschaftsstudie - Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn.* Opladen: Leske + Budrich.

Baumert, J., Brunner, M., Lüdtke, O., & Trautwein, U. (2007). Was messen internationale Schulleistungsstudien? – Resultate kumulativer Wissenserwerbsprozesse. *Psychologische Rundschau, 58*(2), 118-145.

Beck, B., & Klieme, E. (Hrsg.). (2007). *Sprachliche Kompetenzen: Konzepte und Messung: DESI-Studie (Deutsch Englisch Schülerleistungen International)*. Weinheim: Beltz.

Beck, I. L., McKeown, M. G., Sinatra, G. M., & Loxterman, J. A. (1991). Revising social studies text from a text-processing perspective: Evidence of improved comprehensibility. *Reading Research Quarterly, 26*, 251-276.

Best, R. M., Rowe, M., Ozuru, Y., & McNamara, D. S. (2005). Deep-level comprehension of science texts: The role of the reader and the text. *Topics in Language Disorders, 25*, 62-80.

Biologie heute entdecken. (2004). *Allgemeine Ausgabe für SII*. Braunschweig: Schroedel.

Biologie Oberstufe. (2001). *Gesamtband*. Berlin: Cornelsen.

Björnsson, C. H. (1968). *Läsbarhet (deutsch: Lesbarkeit durch Lix, Übersetzung von Studienrätin Helga Johansson, 1968)*. Stockholm: Pedagogiskt Centrum vid Stockholms skoldirektion.

Bloom, B. S. (1974). *Taxonomie von Lernzielen im kognitiven Bereich.* Weinheim: und Basel: Beltz.

Bloom, B. S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain*. New York: David McKay.

Bol, L., & Strage, A. (1996). The contradiction between teachers' instructional goals and their assessment practices in high school biology courses. *Science Education, 80* (2), 145-163.

Bonfadelli, H., Jarren, O., & Siegert, G. (Hrsg.). (2005). *Einführung in die Publizistikwissenschaft*. Bern: Haupt.

Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschafter* (4. überarbeitete Auflage). Berlin: Springer.

Boueke, D. (Hrsg.). (1979). *Deutschunterricht in der Diskussion: Forschungsberichte* (2. Auflage). Paderborn: Schöningh.

Britton, B. K., Gulgoz, S., & Glynn, S. (1993). Impact of good and poor writing on learners: Research and theory. In B. K. Britton, A. Woodward & M. R. Binkley (Eds.), *Learning from textbooks: Theory and practice* (pp. 1-46). Hillsdale: Erlbaum.

Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion* (2. Auflage). München: Pearson Studium.

Bundesamt für Statistik (BfS) (2008a). Abschlüsse der universitären Hochschulen: Basistabellen. Internet: http://www.bfs.admin.ch/bfs/portal/de/index/themen/15/06/data.html. Zugriff am 23. Juni 2008.

Bundesamt für Statistik (BfS) (2008b). Maturitätsquote. Internet: http://www.bfs.admin.ch/bfs/portal/de/index/themen/15/06/dos/blank/05/03.html. Zugriff am 23. Juni 2008.

Bybee, R. (1997). *Achieving scientific literacy: from purposes to practices*. Portsmouth: Heilman.

Campbell, N. A., & Reece, J. B. (2006). *Biology (*6th edition). München: Pearson Education.

Carlsen, W. S. (1991). Questioning in classrooms: A sociolinguistic perspective. *Review of Educational Research, 61*, 157-178.

Cazden, C. B. (1986). Classroom discourse. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 432-463). New York: Macmillan.

Chamorro-Premuzic, T., & Furnham, A. (2004). A possible model for explaining the personality-intelligence interface. *British Journal of Psychology, 95*, 249-264.

Chamorro-Premuzic, T., & Furnham, A. (2008). Personality, intelligence, approaches to learning and academic performance. *Personality and Individual Differences, 44*, 1596-1603.

Chiesi, H. L., Spilich, G. J., & Voss, J. F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning & Verbal Behavior, 18*, 257-273.

Chipman, S. F., Segal, J. W., & Glaser, R. (Eds.). (1985). *Thinking and learning skills, vol.2*. Hillsdale, N.J.: Erlbaum.

Cochran, W. G. (1977). *Sampling Techniques*. New York: John Wiley & Sons.

DESI-Konsortium. (2006). Unterricht und Kompetenzerwerb in Deutsch und Englisch. Zentrale Befunde der Studie Deutsch-Englisch-Schülerleistungen-International (DESI). Frankfurt am Main: Deutsches Institut für Internationale Pädagogische Forschung (DIPF). Internet: http://www.dipf.de/desi/DESI_Zentrale_Befunde.pdf. Zugriff am 14.03.2006.

DESI-Konsortium. (Hrsg.). (2008). *Unterricht und Kompetenzerwerb in Deutsch und Englisch: Ergebnisse der DESI-Studie*. Weinheim: Beltz.

Diakidoy, I. N. (1999). Comprehension and learning from scientific text. In A. Gagatsis (Ed.), *A multidimensional approach to learning in mathematics and science* (pp. 385-413). Nicosia, Cyprus: Intercollege Press.

Dochy, F. J. R. C. (1992). *Assessment of prior knowledge as a determinant for future learning: The use of prior knowledge state tests and knowledge profiles*. London: Lemma B.V.

Dochy, F. J. R. C., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research, 69*, 145-186.

Dresel, M., Stöger, H., & Ziegler, A. (2006). Klassen- und Schulunterschiede im Ausmaß von Geschlechtsunterschieden bei Leistungsbewertungen und Leistungsaspirationen: Ergebnisse einer Mehrebenenanalyse. *Psychologie in Erziehung und Unterricht, 53*, 44-61.

Drinkmann, A., & Groeben, N. (1989). *Metaanalysen für Textwirkungsforschung*. Weinheim: Deutscher Studienverlag.

Eberle, F. (1986). *Unterschiede in schulleistungsrelevanten Merkmalen zwischen Wirtschaftsgymnasiasten und Gymnasiasten anderer Maturitätstypen*. Bamberg: Difo-Druck, 2 Bände.

Eberle, F. (2005). *Definitives Konzept für die Phase II des Projekts „Evaluation der Maturitätsreform MAR 95" (EVAMAR) vom 11. April 2005 zuhanden des Bundesamts für Bildung und Wissenschaft und der Konferenz der kantonalen Erziehungsdirektoren (EDK)* (Internes Dokument). Zürich: Universität Zürich.

Eberle, F. (2008). Ist das Bildungsziel des MAR noch zeitgemäss? *AMV aktuell, 2*, 19-28.

EDK. (1994). *Rahmenlehrplan für die Maturitätsschulen (Dossier 30A)*. Bern: Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK).

EDK. (1995). *Verordnung des Bundesrates/Reglement der EDK über die Anerkennung von gymnasialen Maturitätsausweisen (MAR) vom 16. Januar/15. Februar 1995*. Bern: Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK).

EDK. (2004a). HarmoS – Zielsetzungen und Konzeption. Bern: EDK. Internet: http://www.edudoc.ch/static/web/arbeiten/harmos/weissbuch_d.pdf. Zugriff am 09.07.2008.

EDK. (2004b). Jahrespressekonferenz der EDK: Ziele der Volksschule landesweit und verbindlich vereinheitlichen. Medienmitteilung 01.07.2004. Bern: EDK. Internet: http://www.edk.ch/dyn/13596.php. Zugriff am 09.07.2008.

EDK. (2005). HarmoS: Konsortien. Internet: http://www.edudoc.ch/static/web/arbeiten/harmos/harmos_kons_d.pdf. Zugriff am 09.07.2008.

EDK. (2007a). Interkantonale Vereinbarung über die Harmonisierung der obligatorischen Schule vom 14. Juni 2007. Bern: EDK. Internet: http://www.edk.ch/dyn/11659.php. Zugriff am 09.07.2008.

EDK. (2007b). *Teilrevision der Maturitäts-Anerkennungsverordnung des Bundes und des Schweizerischen Maturitätsanerkennungsreglements der Schweizerischen Konferenz der kantonalen Erziehungsdirektoren.* Bern: EDK.

EDK. (2008). Interkantonale Vereinbarung über die Harmonisierung der obligatorischen Schule (HarmoS-Konkordat) vom 14. Juni 2007: Stand kantonale Beitrittsverfahren. (Stand 2. Juli 2008). Bern: EDK. Internet: http://www.edudoc.ch/static/web/arbeiten/harmos/liste_rat_df.pdf. Zugriff am 09.07.2008.

Ehlich, K. (2003). Universitäre Textarten, universitäre Struktur. In K. Ehlich & A. Steets (Hrsg.), *Wissenschaftlich schreiben - lehren und lernen* (S. 13-28). Berlin: De Gruyter.

Ehlich, K., & Steets, A. (2003). Wissenschaftliche Schreibanforderungen in den Disziplinen. Eine Umfrage unter ProfessorInnen der LMU. In K. Ehlich & A. Steets (Hrsg.), *Wissenschaftlich schreiben - lehren und lernen* (S. 129-154). Berlin: De Gruyter.

Elliot, A. J., & Thrash, T. M. (2001). Achievement goals and the hierarchical model of achievement motivation. *Educational Psychology Review, 13*, 139-156.

Ernst, R. (2004). Verantwortungsbewusste Bildung für die Zukunft: Eine Herausforderung für Mittel- und Hochschule. *Gymnasium Helveticum, 1*, 6-10.

Europarat - Rat für kulturelle Zusammenarbeit (2001). Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen. Berlin: Langenscheidt. Internet: http://www.goethe.de/z/50/commeuro/i0.htm. Zugriff am 27.03.2006.

Feilke, H., & Steinhoff, T. (2003). Zur Modellierung der Entwicklung wissenschaftlicher Schreibfähigkeiten. In K. Ehlich & A. Steets (Hrsg.), *Wissenschaftlich schreiben - lehren und lernen* (S. 112-128). Berlin: De Gruyter.

Fenollar, P., Roman, S., & Cuestas, P. J. (2007). University students' academic performance: An integrative conceptual framework and empirical analysis. *British Journal of Educational Psychology, 77*(4), 873-891.

Friedrich, H. F., & Mandl, H. (1992). Lern- und Denkstrategien - ein Problemaufriss. In H. Mandl & H. F. Friedrich (Hrsg.), *Lern- und Denkstrategien. Analyse und Intervention.* (S. 3-54). Göttingen: Hogrefe.

Früh, W. (2004). *Inhaltsanalyse. Theorie und Praxis*. Stuttgart: UTB.

Fucks, W. (1955). *Mathematische Analyse von Sprachelementen, Sprachstil und Sprachen*. Köln: Westdt. Verlag.

Furnham, A., Chamorro-Premuzic, T., & McDougall, F. (2003). Personality, cognitive ability, and beliefs about intelligence as predictors of academic performance. *Learning and Individual Differences, 14*, 49-66.

Garner, R. (1987). *Metacognition and reading comprehension*. Norwood, NJ: Ablex.

Georghiades, P. (2000). Beyond conceptual change learning in science education: focusing on transfer, durability and metacognition. *Educational Research, 42*(2), 119-139.

Glinz, E., Glinz, H., & Ramseier, M. (1997). *Sprachunterricht - Theorie und Praxis: Grundlagen zum "Schweizer Sprachbuch"*. Zürich: SABE.

Glinz, H. (1987). Die Sprachtheorie in und hinter den Lehrern und die Entwicklung der Sprachfähigkeit in den Schülern. In R. Wimmer (Hrsg.), *Sprachtheorie. Der Sprachbegriff in Wissenschaft und Alltag. Jahrbuch des Instituts für deutsche Sprache (Sprache der Gegenwart 71)*. Düsseldorf: Schwann.

Gräber, W., Nentwig, P., & Nicolson, P. (Hrsg.). (2002). *Der Beitrag der Naturwissenschaften zur Allgemeinen Bildung*. Opladen: Leske + Budrich.

Groeben, N. (1972). *Die Verständlichkeit von Unterrichtstexten: Dimensionen und Kriterien rezeptiver Lernstadien*. Münster: Aschendorff.

Hager, W. (1995). *Programme zur Förderung des Denkens bei Kindern*. Göttingen: Hogrefe.

Haider, G., & Reiter, C. (Hrsg.). (2001). *Pisa 2000: Nationaler Bericht. Internationale und nationale Ergebnisse. Vergleich der Schülerleistungen mit dem Schwerpunkt Lesen und Leseverständnis. Beiträge zur vergleichenden Schulforschung*. Innsbruck: Studien-Verlag.

Haider, G., & Reiter, C. (Hrsg.). (2004). *Pisa 2003 – Internationaler Vergleich von Schülerleistungen. Ergebnisse im Überblick. Executive Summary*. Graz: Universität Salzburg, Abteilung Erziehungswissenschaften.

Hailikari, T., Nevgi, A., & Komulainen, E. (2008). Academic self-beliefs and prior knowledge as predictors of student achievement in mathematics: A structural model. *Educational Psychology Review, 28*(1), 59-71.

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education, 15*(3), 309-334.

Halpern, D. F., & LaMay, M. L. (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review, 12*, 229-246.

Hänsgen, K.-D., & Spicher, B. (2001). *EMS Eignungstest für das Medizinstudium in der Schweiz. Vorhersage des Prüfungserfolges*: Universität Freiburg.

Hänsgen, K.-D., & Spicher, B. (2007). *EMS Eignungstest für das Medizinstudium in der Schweiz 2007*: Universität Freiburg.

Heller, K. A. (Hrsg.). (2002). *Begabtenförderung im Gymnasium. Ergebnisse einer zehnjährigen Längsschnittstudie*. Opladen: Leske & Budrich.

Hornung, A. (2005). Textkompetenz und Sprachkompetenz – einige praxisgeleitete Überlegungen zu einem aktuellen Thema. In Höheres Lehramt Berufsschulen und Höheres Lehramt Mittelschulen am Zürcher Hochschulinstitut für Schulpädagogik und Fachdidaktik (Hrsg.), *Texte verstehen - Berichte aus Forschung und Praxis Bd. 2* (S. 128-135). Bern und Zürich: h.e.p. und Pestalozzianum.

Huber, L. (1997). Fähigkeit zum Studieren – Bildung durch Wissenschaft. Zum Problem der Passung zwischen Gymnasialer Oberstufe und Hochschule. In E. Liebau, W. Mack & C. Scheilke (Hrsg.), *Das Gymnasium. Alltag, Reform, Geschichte, Theorie* (S. 333–351). Weinheim: Juventa.

Huber, L. (1998). Allgemeine Studierfähigkeit, basale Fähigkeiten, Grundbildung. Zur aktuellen Diskussion um die gymnasiale Oberstufe. In R. Messner, E. Wicke & D. Bosse (Hrsg.), *Die Zukunft der gymnasialen Oberstufe* (S. 150-181). Weinheim: Beltz.

Huber, L. (im Druck). Von „basalen Fähigkeiten“ bis „vertiefte Allgemeinbildung“: Was sollen Abiturienten für das Studium mitbringen? In D. Bosse (Hrsg.), *Gymnasiale Bildung zwischen Kompetenzorientierung und Kulturarbeit*. Wiesbaden: Verlag für Sozialwissenschaften.

Ingendahl, W. (1999). *Sprachreflexion statt Grammatik. Ein didaktisches Konzept für alle Schulstufen*. Tübingen: Max Niemeyer.

Jordan, A., Ross, N., Krauss, S., Baumert, J., Blum, W., Neubrand, M., et al. (2006). *Klassifikationsschema für Mathematikaufgaben: Dokumentation der Aufgabenklassifikation im COACTIV-Projekt. Materialien aus der Bildungsforschung, Nr. 81*. Berlin: Max-Planck-Institut für Bildungsforschung.

Kendeou, P., & van den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition, 35*, 1567-1577.

Kim, S., & Van Dusen, L. M. (1998). The role of prior knowledge and elaboration in text comprehension and memory: A comparison of self-generated and text provided elaboration. *American Journal of Psychology, 111*, 353-378.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*, 163-182.

Kintsch, W. (1998). *Comprehension: a paradigm for cognition*. Cambridge: Cambridge University Press.

Kish, L. (1965). *Survey Sampling*. New York: John Wiley & Sons.

Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. *Zeitschrift für Pädagogik, 52*, 876-903.

Klieme, E., Neubrand, M., & Lüdtke, O. (2001). Mathematische Grundbildung: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiss (Hrsg.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske & Budrich.

Klieme, E., Neubrand, M., & Lüdtke, O. (2002). *Kompetenzstufen und Schwierigkeitsmodelle für den PISA-Test zur mathematischen Grundbildung*. Opladen: Leske & Budrich.

KMK. (2004). Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland: Einheitliche Prüfungsanforderungen in der Abiturprüfung Biologie (Beschluss der Kultusministerkonferenz vom 01.12.1989 i.d.F. vom 05.02.2004). Internet: http://www.kmk.org/doc/beschl/EPA-Biologie.pdf. Zugriff am 21.02.08.

KMK. (2005). Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland: KMK-Bildungsstandards im Fach Biologie für den Mittleren Bildungsabschluss vom 16.12.2004. München: Wolters Kluwer. Internet: http://www.kmk.org/schul/Bildungsstandards/Biologie_MSA_16-12-04.pdf. Zugriff am 17.08.08.

Köller, O., & Klieme, E. (2000). Geschlechtsdifferenzen in den mathematisch-naturwissenschaftlichen Leistungen. In J. Baumert, W. Bos & R. Lehmann (Hrsg.), *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie - Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Band 2. Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe* (S. 373-404). Opladen: Leske + Budrich.

Köller, O., Watermann, R., Trautwein, U., & Lüdtke, O. (2004). *Wege zur Hochschulreife in Baden-Württemberg – TOSCA – Eine Untersuchung an allgemein bildenden und beruflichen Gymnasien*. Opladen: Leske + Budrich.

Kommission für Schul- & Gemeindebibliotheken des Kantons Bern (Hrsg.). (1989). Textbeurteilung: eine vielschichtige Angelegenheit. Lesbarkeits-Index (LIX): ein wertvolles Hilfsmittel. *Der Berner Bibliothekar, 46*, 2-12.

Köster, J. (2005). Wodurch wird ein Text schwierig? Ein Test für die Fachkonferenz. *Deutschunterricht, 5*, 34-39.

Kretschmer, I. (1981). Untersuchung zur Förderung des produktiven Denkens von elfjährigen Schülern. *Die deutsche Schule, 73*(4), 205-214.

Kruse, O. (2007). Schreibkompetenz und Studierfähigkeit. Mit welchen Schreibkompetenzen sollten die Schulen ihre Absolvent/innen ins Studium entlassen? In M. Becker-Mrotzek & K. Schindler (Hrsg.), *Texte schreiben. Kölner Beiträge zur Sprachdidaktik (5)* (S. 117-143). Duisburg: Gilles & Francke.

Kubinger, K. D. (2003). Adaptives Testen. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 1-9). Weinheim: Beltz.

Kunter, M., Schümer, G., Artelt, C., Baumert, J., Klieme, E., Neubrand, M., et al. (2002). *Pisa 2000 - Dokumentation der Erhebungsinstrumente*. Berlin: MPI für Bildungsforschung.

Lehtonen, R., & Pahkinen, E. J. (1994). *Practical Methods for Design and Analysis of Complex Surveys*. New York: John Wiley & Sons.

Leutner, D., Wirth, J., & Fleischer, J. (2005). Zentrale Lernstandserhebungen in der Jahrgangsstufe 9 im Jahr 2004 in NRW: Erster Kurzbericht zur wissenschaftlichen Begleitung. Universität Duisburg-Essen. Lehrstuhl für Lehr-Lernpsychologie. Internet: http://www.standardsicherung.nrw.de/lernstand8/upload/download/ergebn_05/kurzbericht1_wb_04.pdf. Zugriff am 27.03.2006.

Levack, B. P. (2003). *Hexenjagd. Die Geschichte der Hexenverfolgungen in Europa*. München: Beck.

Linder Biologie. (2005). *Gesamtband SII (*22. Auflage*)*. Braunschweig: Schroedel.

Linneweber, H., & Wälti, B. (2006). *Zwischenbericht HarmoS Konsortium Mathematik - für die erste Projektetappe: Mai 2005 - Juni 2006, Stand der Arbeiten August 2006*. Aarau: Fachhochschule Nordwestschweiz.

Long, D. L., Seely, M. R., & Oppy, B. J. (1996). The availability of causal information during reading. *Discourse Processes, 22*, 145-170.

Maradan, O., & Mangold, M. (2005). Bildungsstandards in der Schweiz. Das Projekt HarmoS. *PH-akzente, 2*, 3-7.

Marton, F., & Säljö, R. (1976). On qualitative differences in learning II: Outcome as a function of the learner's conception of the task. *British Journal of Educational Psychology, 46*, 115-127.

MAV (1968). *Verordnung über die Anerkennung von Maturitätsausweisen (Maturitäts-Anerkennungsverordnung - MAV) vom 22. Mai 1968. Sammlung der eidgenössischen Gesetze, 1968,* 693-703.

Mayr, E. (2001). *What Evolution Is*. New York: Basic Books.

McCrae, R. R., & Costa, P. T., Jr. (1997). Personality trait structure as a human universal. *American Psychologist, 52*, 509-516.

McNamara, D. S. (2001). Reading both high-coherence and low coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology, 55*, 51-62.

McNamara, D. S., & Kintsch, W. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes, 22*, 247-287.

Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (2005a). Kompetenzniveaus in den 2005 im Rahmen der Lernstandserhebungen 9 getesteten Bereichen. Anforde-

rungsniveaus der Aufgaben und Kompetenzniveaus der Schülerinnen und Schüler. Soest: Landesinstitut für Schule und Weiterbildung. Internet: http://www.standardsicherung.nrw.de/lernstand8/upload/download/ergebn_05/kompetenzniveaus_05.pdf. Zugriff am 27.03.2006.

Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (2005b). Zentrale Lernstandserhebungen in Klasse 9 – Landesweite Ergebnisse 2005. Soest: Landesinstitut für Schule und Weiterbildung. Internet: http://www.standardsicherung.nrw.de/lernstand8/upload/download/ergebn_05/lse-ergebnisse_2005.pdf. Zugriff am 27.03.2006.

Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (2007). Erläuterungen zu den Kompetenzniveaus Deutsch – Leseverstehen bei der Lernstandserhebung 2007 (lernstand 8). Soest: Stab Bildungsforschung und Grundlagen der Standardentwicklung und -überprüfung (StB). Internet: http://www.standardsicherung.schulministerium.nrw.de/lernstand8/upload/download/erg_06-07/komp-niv_de_leseverstehen07.pdf. Zugriff am 16.07.2008.

Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (2008). Zentrale Lernstandserhebungen in der Jahrgangsstufe 8. Soest: Stab Bildungsforschung und Grundlagen der Standardentwicklung und -überprüfung (StB). Internet: http://www.standardsicherung.schulministerium.nrw.de/lernstand8/. Zugriff am 16.07.2008.

MNU (Deutscher Verein zur Förderung des mathematischen und naturwissenschaftlichen Unterrichts e.V.). (2005). *Arbeiten mit den Bildungsstandards im Fach Biologie fachspezifisch und fachübergreifend, dimensioniert und niveauvoll. Empfehlungen für die Umsetzung der KMK-Standards Biologie S I.*

Mokhtari, K., & Reichard, C. A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology, 94*, 249-259.

Mugabushaka, A.-M. (2005). Schlüsselqualifikationen im Hochschulbereich. Eine diskursanalytische Untersuchung der Modelle, Kontexte und Dimensionen in Deutschland und Großbritannien. Dissertation. Online-Publikation. Universität Kassel. Internet: http://kobra.bibliothek.uni-kassel.de/bitstream/urn:nbn:de:hebis:34-2118/1/dis3233_05.pdf. Zugriff am 10.10.07.

Narciss, S., & Proske, A. (2001). *Lern- und Studieraufgaben für die universitäre Lehre*: Technische Universität Dresden (Unveröffentlichtes Manuskript).

NCTM. (2000). *National Council of Teachers of Mathematics: Principles and standards for school mathematics*. Reston, VA: NCTM.

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*, 77-101.

Nold, G. (2007). DESI im Kontext des Gemeinsamen Europäischen Referenzrahmens für Sprachen. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen: Konzepte und Messung: DESI-Studie (Deutsch Englisch Schülerleistungen International)* (S. 299-305). Weinheim: Beltz.

Notter, P., & Arnold, C. (2003). *Der Übergang ins Studium. Bericht zu einem Projekt der Konferenz der Schweizerischen Gymnasialrektoren (KSGR) und der Rektorenkonferenz der Schweizer Universitäten (CRUS).* Bern: Bundesamt für Bildung und Wissenschaft.

Notter, P., & Arnold, C. (2006). *Der Übergang ins Studium II, Bericht zu einem Projekt der Konferenz der Schweizerischen Gymnasialrektoren (KSGR) und der Rektorenkonferenz der Schweizer Universitäten (CRUS)*. Bern: Staatssekretariat für Bildung und Forschung, Schriftenreihe SBF.

O'Connor, M., & Paunonen, S. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences, 43*, 971–990.

O'Reilly, T., & McNamara, D. S. (2007). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional "high-stakes" measures of high school students' science achievement. *American educational research journal, 44* (1), 161-196.

OECD (Ed.). (1999). *Measuring student knowledge and skills. A new framework for assessment*. Paris: Organisation for Economic Co-operation and Development (OECD).

OECD (Ed.). (2007). *PISA 2006 Science Competencies for Tomorrow's World: Vol 1.* Paris: Organisation for Economic Co-operation and Development (OECD).

Oelkers, J. (2006). *Bildungsstandards am Gymnasium: Ein neues Problem?* Vortrag auf der 2. Kadertagung des Instituts Sekundarstufe II der Pädagogischen Hochschule Bern am 30. August 2006 an der Universität Bern.

Oelkers, J. (2008). *Die Qualität der Schweizer Gymnasien*. Bern: h.e.p.

Ossner, J. (2006). Kompetenzen und Kompetenzmodelle im Deutschunterricht. *Didaktik Deutsch, 21*, 5-19.

Ozuru, Y., Dempsey, K., & McNamara, D. S. (in press). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*.

Prenzel, M., Drechsel, B., Carstensen, C. H., & Ramm, G. (2004). PISA 2003 – eine Einführung. In M. Prenzel, J. Baumert, W. Blum, R. H. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003: Der Bildungsstand der Jugendlichen in Deutschland - Ergebnisse des zweiten internationalen Vergleichs* (S. 13-46). Münster: Waxmann.

Purves, W. K. (2008). *Life - The Science of Biology* (8th edition). Sunderland: Sinauer Associates.

Ramseier, E., Allraum, J., Stalder, U., Grin, F., Alliata, R., Müller, S., et al. (2004). *Evaluation der Maturitätsreform 1995 (EVAMAR). Neue Fächerstruktur – Pädagogische Ziele – Schulentwicklung. Schlussbericht zur Phase 1.* Bern: Schweizerische Konferenz der Erziehungsdirektoren und Bundesamt für Bildung und Wissenschaft.

Ramseier, E., Keller, C., & Moser, U. (1999). *Bilanz Bildung. Eine Evaluation am Ende der Sekundarstufe II auf der Grundlage der „Third International Mathematics and Science Study"*. Chur: Rüegger.

Recht, D. R., & Leslie, L. (1988). Effect of prior knowledge on good and poor readers' memory of text. *Journal of Educational Psychology, 80*, 16-20.

Resnick, L. (1987). *Education and learning to think*. Washington, D.C.: National Academy Press.

Richards, E., & Singer, M. (2001). Representation of complex goal structures in narrative comprehension. *Discourse Processes, 31*, 111-135.

Rost, J. (2004). *Lehrbuch Testtheorie Testkonstruktion*. Bern: Huber.

Satin, A., & Shastry, W. (1999). *Survey Sampling*. Ottawa: Statistics Canada.

SBF (2008). Staatssekretariat für Bildung und Forschung. Schweizerische Maturitätsprüfung, Richtlinien für die Jahre 2003-2006, Gültigkeit verlängert bis 2008, Bereich Naturwissenschaften. Internet: http://www.sbf.admin.ch/htm/themen/bildung/matur/ch-matur_de.html. Zugriff am 15.07.08.

Schallberger, U. (1996). Anforderungen an das Zulassungsverfahren in der Schweiz. Leitlinien für die Entwicklung eines eignungsdiagnostischen Verfahrens. In K.-D. Hänsgen & N. Ischi (Hrsg.),

Eignungsprüfung für das Medizinstudium. Kriterien und Testverfahren (S. 92-101). Fribourg: Zentrum für Testentwicklung und Diagnostik.

Schiefele, U., Artelt, C., Schneider, W., & Stanat, P. (Hrsg.). (2004). *Struktur, Entwicklung und Förderung von Lesekompetenz. Vertiefende Analysen im Rahmen von PISA 2000*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Schneider, G., & North, B. (1999). *In anderen Sprachen kann ich…" Skalen zur Beschreibung, Beurteilung und Selbsteinschätzung der fremdsprachlichen Kommunikationsfähigkeit: Umsetzungsbericht. Nationales Forschungsprogramm 33 zur Wirksamkeit unserer Bildungssysteme*. Bern/Aarau: Schweizerische Koordinationsstelle für Bildungsforschung SKBF.

Schüpbach, H., Pixner, J., & Zapf, S. (2006). Handlungskompetenz im Hochschulstudium. *Bildung und Erziehung, 59* (2), 147-166.

Shapiro, A. M. (2004). How including prior knowledge as a subject variable may change outcomes of learning research. *American Educational Research Journal, 41* (1), 159-189.

Sieber, P. (2006a). HarmoS Erstsprache. Kadertagung Bildungsstandards 27. Januar 2006. Unveröffentlichte Präsentation.

Sieber, P. (2006b). Schritte auf dem Weg zu einem Kompetenzmodell Erstsprache. Symposion Deutschdidaktik Weingarten. Zürich: Pädagogische Hochschule. Internet: http://www.symposion-deutschdidaktik.de/cweb/cgi-bin-noauth/cache/VAL_BLOB/163/163/94/11_Sieber.pdf. Zugriff am 09.07.2006.

Sieber, P. (2006c). Schritte auf dem Weg zu einem Kompetenzmodell Erstsprache. Vernehmlassungstext April 2006. Unveröffentlichtes Manuskript.

Sieber, P. (Hrsg.). (1994). *Sprachfähigkeiten - besser als ihr Ruf und nötiger denn je! : Ergebnisse und Folgerungen aus einem Forschungsprojekt*. Aarau: Sauerländer.

Snow, C. P. (1963). *The Two Cultures: A Second Look*. Cambridge: Cambridge UP.

Social Science Consulting. (2007). TextQuest – Software für Textanalysen. Osnabrück. Internet: http://www.textquest.de/lesbar.htm. Zugriff am 21.07.2008.

Stanat, P., & Kunter, M. (2001). Geschlechterunterschiede in Basiskompetenzen. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 249-269). Opladen: Leske + Budrich.

Steiner, G. (2001). *Lernen – 20 Szenarien aus dem Alltag. Eine Einführung in die Lernpsychologie* (3. Aufl.). Bern: Huber.

Streblow, L. (2004). Zur Förderung der Lesekompetenz. In U. Schiefele, C. Artelt, W. Schneider & P. Stanat (Hrsg.), *Struktur, Entwicklung und Förderung von Lesekompetenz. Vertiefende Analysen im Rahmen von PISA 2000* (S. 275-306). Wiesbaden: VS Verlag für Sozialwissenschaften.

Szidat, J. (2003). Erwartungen der Universität an die Mittelschule. *Bulletin der Vereinigung schweizerischer Hochschuldozenten, 2/3*, 14-17.

Tait, H., & Entwistle, N. J. (1996). Identifying students at risk through ineffective study strategies. *Higher Education, 31*, 97-116.

Takala, S. (2007). Relating Examinations to the Common European Framework. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen: Konzepte und Messung: DESI-Studie (Deutsch Englisch Schülerleistungen International)* (S. 306-331). Weinheim: Beltz.

Tauber, M. (1984). *Leserangepasste Verständlichkeit: der Einfluss von Lesbarkeit und Gliederung am Beispiel von Zeitungsartikeln*. Bern: Lang.

Thim-Mabrey, C. (2005). Stilnormen als Textsortennormen. Korrektur und Beratung zu Texten von Schülern und Studierenden. In K. Adamzik & W.-D. Krause (Hrsg.), *Text-Arbeiten. Textsorten im fremd- und muttersprachlichen Unterricht an Schule und Hochschule* (S. 31-43). Tübingen: Gunter Narr.

Thompson, R. A., & Zamboanga, B. L. (2004). Academic aptitude and prior knowledge as predictors of student achievement in introduction to psychology. *Journal of Educational Psychology, 96*, 778-784.

Tobias, S. (1994). Interest, prior knowledge, and learning. *Review of Educational Research, 64* (1), 37-54.

Trappmann, S., Hell, B., Weigand, S., & Schuler, H. (2007). Die Validität von Schulnoten zur Vorhersage des Studienerfolgs – eine Metaanalyse. *Zeitschrift für Pädagogische Psychologie, 21*(1), 11-27.

Trautwein, U., Watermann, R., Lüdtke, O., Köller, O., Baumert, J., Maaz, K., et al. (2006). *Transformation des Sekundarschulsystems und akademische Karrieren (TOSCA). Dokumentation der eingesetzten Items und Skalen*. Berlin: Max - Planck - Institut für Bildungsforschung.

Trost, G. (1989). A nationwide testing program for admission to medical schools in West Germany. In R. C. King & J. K. Collins (Eds.), *Social applications and issues in psychology* (pp. 131-137). Amsterdam: Elsevier Science.

Tzeng, Y., van den Broek, P., Kendeou, P., & Lee, C. (2005). The computational implementation of the Landscape Model: Modeling inferential processes and memory representations of text comprehension. *Behavioral Research Methods, 37*, 277-286.

van den Broek, P., & Kendeou, P. (2008). Cognitive processes in comprehension of science texts: The role of co-activation in confronting misconceptions. *Applied Cognitive Psychology, 22*, 335-351.

van den Broek, P., Kendeou, P., Sung, Y. C., & Chen, M. (2003). *Comprehension and memory of science texts: A simulation using the Landscape Model*. Paper presented at the Annual Meeting of the Society for Text and Discourse. Madrid, June 2003.

van den Broek, P., Rapp, D. N., & Kendeou, P. (2005). Integrating memory- based and constructionist approaches in accounts of reading comprehension. *Discourse Processes, 39*, 299-316.

van den Broek, P., Risden, K., Fletcher, C. R., & Thurlow, R. (1996). A "landscape" view of reading: Fluctuating patterns of activation and the construction of a stable memory representation. In B. K. Britton & A. C. Graesser (Eds.), *Models of understanding text* (pp. 165-187). Hillsdale, NJ: Lawrence Erlbaum Associates.

van den Broek, P., Risden, K., & Husebye-Hartmann, E. (1995). The role of readers' standards of coherence in the generation of inferences during reading. In R. F. Lorch, Jr. & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 353-373). Hillsdale, NJ: Erlbaum.

van den Broek, P., Virtue, S., Everson, M., Tzeng, Y., & Sung, Y. C. (2002). Comprehension and memory of science texts: Inferential processes and the construction of a mental representation. In J. Otero, J. A. Leon & A. C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 131-154). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71-98). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Vosniadou, S., & Brewer, W. F. (1994). Mental models of the day/night cycle. *Cognitive Science, 18*, 123-183.

Voss, J., & Silfies, L. (1996). Learning from history text: The interaction of knowledge and comprehension skill with text structure. *Cognition and Instruction, 14*, 45-68.

Weinert, F. E. (2001). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessung in Schulen* (S. 17-31). Weinheim: Beltz.

West, L. H. T., & Pines, A. L. (1985). *Cognitive structure and conceptual change*. Orlando, FL: Academic Press.

Wetzenstein, E. (2004). *Entwicklung eines Anforderungsprofils für Studierende am Beispiel der Psychologie*. Beitrag anlässlich des Bühler-Kolloquium der TU Dresden.

Willenberg, H. (Hrsg.). (2007). *Kompetenzhandbuch für den Deutschunterricht: auf der empirischen Basis des DESI-Projekts*. Baltmannsweiler: Schneider Hohengehren.

Wilson, P. T., & Anderson, R. C. (1986). What they don't know will hurt them: The role of prior knowledge in comprehension. In J. Orasanu (Ed.), *Reading comprehension: From research to practice* (pp. 31-48). Hillsdale, NJ: Lawrence Erlbaum.

Wonderlic Personnel Test. (1992). *Wonderlic Personnel Test and Scholastic Level Exam: User's Manual*. Libertyville, IL: Wonderlic Personnel Test, Inc.

Wright, B. D., & Stone, M. H. (1979). *Best Test Design. Rasch Measurement*. University of Chicago: Mesa Press.

Zheng, A. Y., Lawhorn, J. K., Lumley, T., & Freeman, S. (2008). Application of Bloom's Taxonomy Debunks the "MCAT Myth". *Science, 319* (5862), 414-415.

Zimmer, K. (2006). PISA 2003 - Kompetenzmuster von Jungen und Mädchen in den deutschen Ländern. *Unterrichtswissenschaft 34*(4), 310-329.

Abbildungs- und Tabellenverzeichnis

Abbildungsverzeichnis

Abb. I.1:	Umfassendes Modell zur Herleitung der Evaluationskriterien	28
Abb. I.2:	Für EVAMAR II ausgewählte Untersuchungen	29
Abb. I.3:	Organigramm von EVAMAR II	33
Abb. II.1:	Zusammenhang zwischen Studienfächern und Unterrichtsfächern	44
Abb. III.1:	Schichtenmodell der Sprachfähigkeiten von Glinz (1987, zitiert nach Sieber, 1994, S. 41)	66
Abb. III.2:	Eigene Darstellung von drei der sechs DESI-Kompetenzbereiche für Deutsch (DESI-Konsortium, 2006) mit zugehörigen Niveaus A-E (ohne Argumentation, Wortschatz, Rechtschreiben)	67
Abb. III.3:	Das PISA-Lesekompetenzmodell mit fünf Kompetenzstufen. Eigene Darstellung der drei Lese-Teilkompetenzen (nach Artelt et al., 2004, S. 144-145).	69
Abb. III.4:	Eigene Darstellung: Das PISA-Lesemodell unter Einbezug der verwendeten Textsorten	71
Abb. III.5:	Gemeinsame Referenzniveaus: Globalskala (Auszug Niveau C) des GER (Europarat, 2001, S. 35)	73
Abb. III.6:	DIALANG-Items zur Selbsteinschätzung – Auszug Niveau C (Europarat, 2001, S. 223-226)	74
Abb. III.7:	HarmoS: Ein Modell für die Beschreibung (Sieber, 2006a, S. 18)	77
Abb. III.8:	HarmoS – ein Beispiel: Hören (Sieber, 2006a, S. 19)	77
Abb. III.9:	Die höchsten Kompetenzniveaus der Lernstandserhebungen Nordrhein-Westfalens 2005 – nachgebildete Matrix	79
Abb. III.10:	Grundmodell Erstsprache nach GER (die von EVAMAR II untersuchten Bereiche sind schraffiert)	82
Abb. III.11:	Kompetenzraster für die Tests in Erstsprache (mit Kann-Beschreibungen)	85
Abb. III.12:	Mögliche Dimensionen eines Kompetenzmodells	86
Abb. III.13:	Darstellungsform Matrix für die Inhalts- und Fähigkeitsdimensionen	86
Abb. III.14:	Das Kompetenzmodell der deutschen Kultusministerkonferenz	87
Abb. III.15:	Kompetenzaspekte und -bereiche bei HarmoS (9. Jahrgangsstufe)	89
Abb. III.16:	Die Matrix der Verhaltenserwartungen und der Inhalte	91
Abb. III.17:	Kompetenzraster Mathematik für EVAMAR II	94
Abb. III.18:	Vergleich der Kapiteleinteilung von drei deutschsprachigen Standardlehrwerken der Biologie auf der Gymnasialstufe (Stufe SII)	98
Abb. III.19:	Vergleich der Kapiteleinteilung zweier übersetzter amerikanischer Standardlehrwerke der Biologie für Studienanfänger (Tertiärstufe)	99
Abb. III.20:	Kurzformeln für die nachfolgend besprochenen Dokumente	100
Abb. III.21:	Einteilung der Biologie in Teilgebiete, angelehnt an deutsch- und englischsprachige Standardbiologielehrwerke, an die deutschen EPA Biologie und an die Regensburger Verbundklassifikation	104
Abb. III.22:	Kategorien naturwissenschaftlicher Bildung (Scientific Literacy) nach Bybee (1997). Übersetzung nach Gräber, W. et al. (Hrsg.) (2002): Scientific Literacy. Leske und Budrich, VS Verlag für Sozialwissenschaften/GWV Fachverlage GmbH, Wiesbaden. S. 21 – 43.	108
Abb. III.23:	Der Kompetenzraster Fachwissen Biologie für die Verwendung in EVAMAR II	111
Abb. IV.1:	Studierstrategien, Leistungsziele und Studienerfolg	117
Abb. IV.2:	Auszug von codierten Sinneinheiten (Fremdwörter)	124
Abb. IV.3:	Wortschatz-Aufgabe zum Originaltext in Abb. IV.4 und zu den codierten Sinneinheiten in Abb. IV.2	125

Abb. IV.4: Originaltext mit den codierten Sinneinheiten, den Niveau-Bestimmungen und der Textauswahl für die Aufgaben (Steiner, 2001, S. 16–17) 126
Abb. IV.5: Detailliertes Leseverstehen – Aufgabe zum Originaltext in Abb. IV.4 und zu den codierten Sinneinheiten in Abb. IV.2 127
Abb. IV.6: Ausschnitt aus einem Skript (Quelle: ETH Zürich, Physik, Felder, Geometrie, S. 30) 130
Abb. IV.7: Testaufgabe zum Lehrtext in Abb. IV.6 131
Abb. IV.8: Seite aus einem Lehrtext im Studienfach Publizistik und Kommunikationswissenschaft 133
Abb. IV.9: Testaufgabe zum Skriptausschnitt in Abb. IV.8 134
Abb. IV.10: Seite aus einem Lehrtext im Studienfach Biologie 135
Abb. IV.11: Testaufgabe zum Lehrtext in Abb. IV.10 136
Abb. IV.12: Item Characteristic Curve 144
Abb. IV.13: Ergebnisse Personenfähigkeiten Erstsprache Total 146
Abb. IV.14: Beispiel einer leichten Aufgabe zum Wortschatz 147
Abb. IV.16: Beispiel einer leichten Grammatikaufgabe 149
Abb. IV.17: Beispiel einer schwierigen Textverständnis-Aufgabe (allgemeines Leseverständnis) 150
Abb. IV.18: Beispiel einer sehr schwierigen Textverständnis-Aufgabe (allgemeines Leseverständnis) im offenen Format (fünfstufige Partial-Credit-Codierung) 152
Abb. IV.19: Beispiel einer schwierigen Textverständnis-Aufgabe (allgemeines Leseverständnis) im offenen Format (vierstufige Partial-Credit-Codierung) 153
Abb. IV.20: Beispiel einer mittelschweren Wortschatz-Aufgabe 154
Abb. IV.21: Ergebnisse Personenfähigkeiten Mathematik Total 155
Abb. IV.22: Beispiel 1 einer gut gelösten Mathematikaufgabe 156
Abb. IV.23: Beispiel 2 einer gut gelösten Mathematikaufgabe 157
Abb. IV.24: Beispiel 3 einer gut gelösten Mathematikaufgabe 158
Abb. IV.25: Beispiel 4 einer gut gelösten Mathematikaufgabe 158
Abb. IV.26: Beispiel 5 einer gut gelösten Mathematikaufgabe 159
Abb. IV.27: Beispiel 1 einer schlecht gelösten Mathematikaufgabe 159
Abb. IV.28: Beispiel 2 einer schlecht gelösten Mathematikaufgabe 159
Abb. IV.29: Beispiel 3 einer schlecht gelösten Mathematikaufgabe 160
Abb. IV.30: Beispiel einer Aufgabe mit unterschiedlich schwierigen Teilaufgaben 160
Abb. IV.31: Ergebnisse Personenfähigkeiten Biologie Total 162
Abb. IV.32: Beispiel 1 eines leichten Items in Biologie 162
Abb. IV.33: Beispiel 2 eines leichten Items in Biologie 163
Abb. IV.34: Beispiel 3 eines leichten Items in Biologie 163
Abb. IV.35: Beispiel 4 eines leichten Items in Biologie 164
Abb. IV.36: Beispiel 1 eines mittelschwierigen Items in Biologie 164
Abb. IV.37: Beispiel 2 eines mittelschwierigen Items in Biologie 165
Abb. IV.38: Beispiel 3 eines mittelschwierigen Items in Biologie 166
Abb. IV.39: Beispiel 4 eines mittelschwierigen Items in Biologie 166
Abb. IV.40: Beispiel 1 eines schwierigen Items in Biologie 167
Abb. IV.41: Beispiel 2 eines schwierigen Items in Biologie 168
Abb. IV.42: Beispiel 3 eines schwierigen Items in Biologie 168
Abb. IV.43: Beispiel 4 eines schwierigen Items in Biologie 169
Abb. IV.44: Ergebnisse Personenfähigkeiten UFT Total 172
Abb. V.1: Anteile der verschiedenen Kategorien an der Gesamtpunktzahl N = 65 (54 Maturaprüfungen, 9 Vor-Maturaprüfungen, 2 Ergänzungswahlfachprüfungen; D 58, F 4, I 3) 267
Abb. V.2: Anteile von Erinnerungs- und Transferaufgaben (Biologie, Gymnasium) 267
Abb. V.3: Vergleich der kognitiven Anforderungen von Maturitätsprüfungen und universitären Prüfungen in Biologie 270
Abb. V.4: Anteile von Erinnerungs- und Transferaufgaben (Biologie, Hochschule) 271

Abb. V.5: Verteilung der kognitiven Aktivierungspotenziale der Aufgaben aus schweizerischen Maturitätsprüfungen 2007 (schräg schraffiert) mit denjenigen Test-Items aus der Item-Selektion Biologie EVAMAR II (waagrecht schraffiert) 272

Abbidungen VI.x: siehe Kapitel VI

Abb. VII.1: Für EVAMAR II ausgewählte Untersuchungen 367
Abb. VII.2: Zusammenhang zwischen Studienfächern und Unterrichtsfächern 369

Tabellenverzeichnis

Tab. II.1: ausgewählte Studienfächer der Datenbasis von TP A 36
Tab. II.2: Vorgegebenes Raster zur Erfassung der Wissensdimension und der Schulfächer 38
Tab. II.3: Wissenskategorien der Sinneinheiten (Hauptkategorien) 39
Tab. II.4: Kategorien kognitiver Fähigkeiten für Aufgaben und Prüfungen 40
Tab. II.5: Bedeutung gymnasialer Fachbereiche für die eigene Lehrveranstaltung 47
Tab. II.6: Bedeutung gymnasialer Fachbereiche nach Studienfächern (Rangzahl jeweils in Klammer) 48
Tab. II.7: Konkret genanntes, vorausgesetztes Fachwissen und -können 50
Tab. II.8: Konkret genannte Lücken beim vorausgesetzten Fachwissen und Können 53
Tab. II.9: Wichtigkeit überfachlicher Kompetenzen für die eigene Lehrveranstaltung 56
Tab. II.10: Defizite in als wichtig erachtetem überfachlichem Wissen und Können 57
Tab. II.11: Weiteres, wünschbares Wissen und Können bei Studienbeginn 58
Tab. IV.1: Wichtige Prädiktoren hoher Studierfähigkeit 118
Tab. IV.2: Die sieben Testhefte in Biologie 121
Tab. IV.3: Grundgesamtheit 141
Tab. IV.4: Stichprobe 141
Tab. IV.5: Anzahl teilnehmende Klassen sowie Maturandinnen und Maturanden nach Straten 142
Tab. IV.6: Gesamtergebnisse Erstsprache 145
Tab. IV.7: Gesamtergebnisse Mathematik 154
Tab. IV.8: Gesamtergebnisse Biologie 161
Tab. IV.9: ausgewählte Maturanoten 170
Tab. IV.10: Gesamtergebnisse Überfachlicher Fähigkeitstest (UFT) 171
Tab. IV.11: Mittelwerte der Skalen der überfachlichen Kompetenzen 173
Tab. IV.12: Mittelwerte der Fragen zur Ausdauer (1 = sehr selten; 5 = sehr oft) 174
Tab. IV.13: Interesse am Unterrichtsfach 174
Tab. IV.14: Vergleich der Ergebnisse in Erstsprache zwischen den Straten: Totalergebnis (TOT_SPRACHE) 178
Tab. IV.15: Vergleich der Ergebnisse in Erstsprache zwischen den Straten: Bereich „Allgemeines Leseverstehen“ (ALLG_LV) 178
Tab. IV.16: Vergleich der Ergebnisse in Erstsprache zwischen den Straten: Bereich „Detailliertes Leseverstehen“ (DETAIL_LV) 178
Tab. IV.17: Vergleich der Ergebnisse in Erstsprache zwischen den Straten: Bereich „Sprachreflexion/Wortschatz“ (WORTS) 179
Tab. IV.18: Vergleich der Ergebnisse in Erstsprache zwischen den Straten: Bereich „Sprachreflexion/Grammatik“ (GRAM) 179
Tab. IV.19: Vergleich der Ergebnisse in Mathematik zwischen den Straten: Totalergebnis (TOT_MATH) 180
Tab. IV.20: Vergleich der Ergebnisse in Mathematik zwischen den Straten: Bereich „Analytisches“ (ANALYSIS) 180

Tab. IV.21: Vergleich der Ergebnisse in Mathematik zwischen den Straten: Bereich „Geometrisches“ (GEOMETRIE) 180
Tab. IV.22: Vergleich der Ergebnisse in Mathematik zwischen den Straten: Bereich „Stochastisches“ (STOCHASTIK) 181
Tab. IV.23: Vergleich der Ergebnisse in Mathematik zwischen den Straten: Bereich „Diverses“ (VERSCH) 181
Tab. IV.24: Vergleich der Ergebnisse in Biologie zwischen den Straten: Totalergebnis (TOT_BIO) 182
Tab. IV.25: Vergleich der Ergebnisse in Biologie zwischen den Straten: Bereich „Evolution“ (B_EVO) 182
Tab. IV.26: Vergleich der Ergebnisse in Biologie zwischen den Straten: Bereich „Struktur und Funktion“ (B_STRU) 182
Tab. IV.27: Vergleich der Ergebnisse in Biologie zwischen den Straten: Bereich „Genetik“ (B_GEN) 183
Tab. IV.28: Vergleich der Ergebnisse in Biologie zwischen den Straten: Bereich „Informationsverarbeitung“ (B_INFO) 183
Tab. IV.29: Vergleich der Ergebnisse in Biologie zwischen den Straten: Bereich „Ökologie“ (B_OEKO) 183
Tab. IV.30: Vergleich der Ergebnisse in Biologie zwischen den Straten: Bereich „Stoffwechsel“ (B_STOFF) 184
Tab. IV.31: Vergleich der Ergebnisse jener Personen in Biologie, die das Fach bereits abgeschlossen hatten, zwischen den Straten: Totalergebnis (TOT_BIO) 184
Tab. IV.32: Vergleich der Ergebnisse jener Personen in Biologie, die das Fach nicht abgeschlossen hatten, zwischen den Straten: Totalergebnis (TOT_BIO) 185
Tab. IV.33: Vergleich der Matura-Gesamtnote in Erstsprache zwischen den Straten 185
Tab. IV.34: Vergleich der Matura-Gesamtnote in Mathematik zwischen den Straten 186
Tab. IV.35: Vergleich der Maturanote in Biologie (Erfahrungsnote) zwischen den Straten 186
Tab. IV.36: Vergleich der Matura-Gesamtnote in Naturwissenschaften zwischen den Straten 186
Tab. IV.37: Vergleich der Ergebnisse des überfachlichen Fähigkeitstests zwischen den Straten: Bereich „Quantitative und formale Probleme“ 187
Tab. IV.38: Vergleich der Ergebnisse des überfachlichen Fähigkeitstests zwischen den Straten: Bereich „Diagramme und Tabellen“ 187
Tab. IV.39: Vergleich der Ergebnisse des überfachlichen Fähigkeitstests zwischen den Straten: Bereich „Textverständnis“ 188
Tab. IV.40: Vergleich der Ergebnisse des überfachlichen Fähigkeitstests zwischen den Straten: Bereich „Naturwissen- schaftliches Grundverständnis“ 188
Tab. IV.41: Vergleich der Befragungsergebnisse „überfachliche Kompetenzen“ zwischen den Straten: Bereich „Zeitplanung“ (1 = schlechte Zeitplanung, 5 = sehr gute Zeitplanung) 189
Tab. IV.42: Vergleich der Befragungsergebnisse „überfachliche Kompetenzen“ zwischen den Straten: Bereich „Prüfungs- vorbereitung“ (1 = ineffiziente Prüfungsvorbereitung, 5 = sehr zeiteffiziente Prüfungsvorbereitung) 189
Tab. IV.43: Vergleich der Befragungsergebnisse „überfachliche Kompetenzen“ zwischen den Straten: Bereich „Zeiteffizienz“ (1 = sehr schlechte Zeiteffizienz, 5 = sehr gute Zeiteffizienz) 189
Tab. IV.44: Vergleich der Ergebnisse des Tests in Erstsprache zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Gesamtergebnis (TOT_SPRACHE) 190
Tab. IV.45: Vergleich der Ergebnisse des Tests in Erstsprache zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Allgemeines Leseverstehen“ (ALLG_LV) 191
Tab. IV.46: Vergleich der Ergebnisse des Tests in Erstsprache zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Detailliertes Leseverstehen“ (DETAIL_LV) 191
Tab. IV.47: Vergleich der Ergebnisse des Tests in Erstsprache zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Sprachreflexion/Wortschatz“ (WORTS) 192
Tab. IV.48: Vergleich der Ergebnisse des Tests in Erstsprache zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Sprachreflexion/Grammatik“ (GRAM) 192

Tab. IV.49: Vergleich der Ergebnisse des Tests in Mathematik zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Gesamtergebnis (TOT_MATH) 193
Tab. IV.50: Vergleich der Ergebnisse des Tests in Mathematik zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Analytisches“ (ANALYSIS) 193
Tab. IV.51: Vergleich der Ergebnisse des Tests in Mathematik zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Geometrisches“ (GEOMETRIE) 194
Tab. IV.52: Vergleich der Ergebnisse des Tests in Mathematik zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Stochastisches“ (STOCHASTIK) 194
Tab. IV.53: Vergleich der Ergebnisse des Tests in Mathematik zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Diverses“ (VERSCH) 195
Tab. IV.54: Vergleich der Ergebnisse des Tests in Biologie zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Totalergebnis (TOT_BIO) 196
Tab. IV.55: Vergleich der Ergebnisse des Tests in Biologie zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Evolution“ (B_EVO) 196
Tab. IV.56: Vergleich der Ergebnisse des Tests in Biologie zwischen den Gruppen unterschiedlicher Schwerpunktfächer Bereich „Struktur und Funktion“ (B_STRU) 197
Tab. IV.57: Vergleich der Ergebnisse des Tests in Biologie zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Genetik“ (B_GEN) 197
Tab. IV.58: Vergleich der Ergebnisse des Tests in Biologie zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Informationsverarbeitung“ (B_INFO) 197
Tab. IV.59: Vergleich der Ergebnisse des Tests in Biologie zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Ökologie“ (B_OEKO) 198
Tab. IV.60: Vergleich der Ergebnisse des Tests in Biologie zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Stoffwechsel“ (B_STOFF) 198
Tab. IV.61: Vergleich der vergebenen Maturanote in Erstsprache zwischen den Gruppen verschiedener Schwerpunktfächer 199
Tab. IV.62: Vergleich der vergebenen Maturanote in Mathematik zwischen den Gruppen verschiedener Schwerpunktfächer 199
Tab. IV.63: Vergleich der vergebenen Matura-Erfahrungsnote in Biologie zwischen den Gruppen verschiedener Schwerpunktfächer 200
Tab. IV.64: Vergleich der vergebenen Maturanote in Naturwissenschaften zwischen den Gruppen verschiedener Schwerpunktfächer 200
Tab. IV.65: Vergleich des Prozentanteils ungenügender Ergebnisse in den Maturitätsprüfungen zwischen den Gruppen unterschiedlicher Schwerpunktfächer 201
Tab. IV.66: Vergleich der Ergebnisse des überfachlichen Fähigkeitstests zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Quantitative und formale Probleme 201
Tab. IV.67: Vergleich der Ergebnisse des überfachlichen Fähigkeitstests zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Diagramme und Tabellen 202
Tab. IV.68: Vergleich der Ergebnisse des überfachlichen Fähigkeitstests zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Textverständnis 202
Tab. IV.69: Vergleich der Ergebnisse des überfachlichen Fähigkeitstests zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Naturwissenschaftliches Grundverständnis 202
Tab. IV.70: Vergleich der Befragungsergebnisse „überfachliche Kompetenzen“ zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Zeitplanung“ (1 = schlechte Zeitplanung, 5 = sehr gute Zeitplanung) 203
Tab. IV.71: Vergleich der Befragungsergebnisse „überfachliche Kompetenzen“ zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Prüfungsvorbereitung“ (1 = ineffiziente Prüfungsvorbereitung, 5 = sehr zeiteffiziente Prüfungsvorbereitung) 203

Tab. IV.72: Vergleich der Befragungsergebnisse „überfachliche Kompetenzen“ zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Zeiteffizienz“ (1 = sehr schlechte Zeiteffizienz, 5 = sehr gute Zeiteffizienz) 204
Tab. IV.73: Vergleich der Befragungsergebnisse „überfachliche Kompetenzen“ zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „vernetztes Denken“ (1 = sehr wenig vernetztes Denken, 5 = sehr viel vernetztes Denken) 204
Tab. IV.74: Vergleich der Befragungsergebnisse „überfachliche Kompetenzen“ zwischen den Gruppen unterschiedlicher Schwerpunktfächer: Bereich „Perspektivenübernahme“ (1 = sehr gute Fremdperspektiven-Übernahme bei Konflikten, 4 = sehr schlechte Fremdperspektiven-Übernahme bei Konflikten) 204
Tab. IV.75: Vergleich Testergebnisse und Maturanoten zwischen den Geschlechtern in Erstsprache 205
Tab. IV.76: Vergleich Testergebnisse und Maturanoten zwischen den Geschlechtern in Mathematik 206
Tab. IV.77: Vergleich Testergebnisse zwischen den Geschlechtern in Biologie und Maturanoten in Naturwissenschaften 207
Tab. IV.78: Vergleich Testergebnisse zwischen den Geschlechtern im UFT 208
Tab. IV.79: Signifikante Unterschiede zwischen den Geschlechtern mit Effektstärke > .2 bei überfachlichen Kompetenzen 208
Tab. IV.80: Darstellung von Unterschieden zwischen Kantonsgruppen mit unterschiedlichen Maturitätsquoten, Erstsprache 209
Tab. IV.81: Darstellung von Unterschieden zwischen Kantonsgruppen mit unterschiedlichen Maturitätsquoten, Mathematik 209
Tab. IV.82: Darstellung von Unterschieden zwischen Kantonsgruppen mit unterschiedlichen Maturitätsquoten, Biologie und Naturwissenschaften 210
Tab. IV.83: Darstellung von Unterschieden zwischen Kantonsgruppen mit unterschiedlichen Maturitätsquoten, UFT 210
Tab. IV.84: Unterschiede zwischen Lang- und Kurzzeitgymnasien: Erstsprache 211
Tab. IV.85: Unterschiede zwischen Lang- und Kurzzeitgymnasien: Mathematik 212
Tab. IV.86: Unterschiede zwischen Lang- und Kurzzeitgymnasien: Biologie und Naturwissenschaften 212
Tab. IV.87: Unterschiede zwischen Lang- und Kurzzeitgymnasien: UFT 213
Tab. IV.88: Vergleich der Ergebnisse in Mathematik nach Sorgfalt der Testbearbeitung 214
Tab. IV.89: Vergleich der Erfahrungsnote in Mathematik nach Sorgfalt der Testbearbeitung 214
Tab. IV.90: Vergleich der schriftlichen Maturaprüfung in Mathematik nach Sorgfalt der Testbearbeitung 214
Tab. IV.91: Vergleich der Maturanoten in Mathematik nach Sorgfalt der Testbearbeitung 215
Tab. IV.92: Mittelwertsvergleich von mündlichen und schriftlichen Noten 215
Tab. IV.93: Korrelative Zusammenhänge zwischen den Tests 216
Tab. IV.94: Korrelationen zwischen dem Test und den Maturanoten in Erstsprache 217
Tab. IV.95: Korrelationen zwischen dem Test und den Maturanoten in Mathematik 217
Tab. IV.96: Korrelationen zwischen dem Test und den Maturanoten in Biologie 217
Tab. IV.97: Korrelationen zwischen dem UFT und den Maturanoten in Erstsprache 218
Tab. IV.98: Korrelationen der Maturanoten untereinander 218
Tab. V.1: Übersicht über die Haupt- und Unterkategorien kognitiver Prozesse nach Anderson et al. (2001) und die im Rahmen von EVAMAR II adaptierten Definitionen der Kategorien (verdeutlicht zunächst nur für das Fach Biologie). 224
Tab. V.2: Die Matura-Aufsatzformen an den Deutschschweizer Gymnasien 228
Tab. V.3: Die Matura-Aufsatzformen in der italienischsprachigen Schweiz 229
Tab. V.4: Die Matura-Aufsatzformen in der französischsprachigen Schweiz 230
Tab. V.5: Die Textsorten der Vorlagen der Deutschschweizer Maturitätsprüfungen 231
Tab. V.6: Die Textsorten der Maturitätsprüfungen aus der Romandie 232
Tab. V.7: Die Textsorten der Maturitätsprüfungen aus der italienischsprachigen Schweiz (Tessin) 233

Tab. V.8: Angaben, aus welchem Jahrhundert die Vorlagen stammen (Schweizer Maturitätsprüfungen, alle drei Sprachregionen) 234
Tab. V.9: Zitierte Autorinnen und Autoren (Auszug) in Deutschschweizer Maturitätsprüfungen 235
Tab. V.10: Herkunft/Nationalität der Autorinnen und Autoren (Deutschschweizer Maturitätsprüfungen) 236
Tab. V.11: Herkunft/Nationalität der Autorinnen und Autoren (Französischsprachige Maturitätsprüfungen aus der Romandie) 237
Tab. V.12: Gender/Geschlecht der Autorinnen und Autoren (Vorlagen an Schweizer Maturitätsprüfungen) 237
Tab. V.13: Länge der Vorlagentexte (Schriftliche Maturitätsprüfungen, alle drei Sprachregionen) 239
Tab. V.14: Quantität der Vorlagentexte (Schweizer Maturitätsprüfungen, alle drei Sprachregionen) 239
Tab. V.15: Auswahlmöglichkeiten an Aufsatzthemen (Maturitätsprüfungen Deutschschweiz und Romandie) 240
Tab. V.16: Erläuternde Hinweise der Lehrperson (Maturitätsprüfungen der Deutschschweiz) 241
Tab. V.17: Erlaubte Hilfsmittel (Maturitätsprüfungen Romandie) (Mehrfachnennungen möglich) 241
Tab. V.18: Analyse der kognitiven Anforderungen (Maturaprüfungen in der Deutschschweiz, Hauptkategorien) 243
Tab. V.19: Analyse der kognitiven Anforderungen (Haupt- und Unterkategorien, gewichtet; Doppelbenennung) (Maturitätsprüfungen Deutschschweiz) 244
Tab. V.20: Analyse der kognitiven Anforderungen (Maturitätsprüfungen im Tessin) 245
Tab. V.21: Analyse der kognitiven Anforderungen (Maturitätsprüfungen in der Romandie) 245
Tab. V.22: Modell der kognitiven Aktivierungspotenziale für Mathematik 249
Tab. V.23: Prozentuale Verteilung der Aufgaben auf die verschiedenen Inhaltsbereiche 251
Tab. V.24: Prozentuale Verteilung der Aufgaben auf die zusammengefassten Inhaltsbereiche 252
Tab. V.25: Kognitive Anforderungen der Aufgabenstellungen 253
Tab. V.26: Quantitative Darstellung des kognitiven Aktivierungspotenzials 257
Tab. V.27: Vergleich der Anforderungen Dozierender mit den Maturitätsprüfungen 258
Tab. V.28: Inhaltsbereiche für die Analyse der Maturaaufgaben 261
Tab. V.29: Durchschnittliches Gewicht des Teilbereichs *t* auf Note oder Gesamtpunktzahl 263
Tab. V.30: Anteil der Gesamtpunktzahl, der durch das Lösen von Aufgaben einer gewissen Kategorie in einer durchschnittlichen Biologie-Maturitätsprüfung erreicht werden kann. 265
Tab. V.31: Vereilung der Gesamtpunktzahl von Hochschulprüfungen in Biologie auf die Kategorien des kognitiven Aktivierungspotenzials 268
Tab. V.32: Anteil der Items innerhalb der Item-Selektion Biologie EVAMAR II mit einem bestimmten kognitiven Aktivierungspotenzial 272
Tab. V.33: Festgestellte Lücken in Biologie und gewünschtes zusätzliches Eingangswissen und -können (Ökologie und Systematik/Evolution kursiv) 273

Tabellen VI.x: siehe Kapitel VI

Abkürzungsverzeichnis

Parameter und Begriffe der Raschmodellierung

θ	Personenfähigkeit
δ	Itemschwierigkeit
ICC	Item Characteristic Curve
IRC	Item Response Curve
p	Lösungswahrscheinlichkeit

Parameter Multiple Mittelwertvergleiche

d	Effektstärke
d max	Effektstärke für die grösste Mittelwertsdifferenz

Fachbereiche Teilprojekt A1

B	Fachbereich Biologie
BG	Fachbereich Bildnerisches Gestalten
C	Computer- und Informatikwissen
Ch	Fachbereich Chemie
D	Fachbereich Deutsch
E	Fachbereich Englisch
F	Fachbereich Französisch
G	Fachbereich Geschichte
Gg	Fachbereich Geographie
Gr	Fachbereich Griechisch
I	Fachbereich Italienisch
L	Fachbereich Latein
M	Fachbereich Mathematik
Mu	Fachbereich Musik
Ph	Fachbereich Physik
PPP	Fachbereich Philosophie, Pädagogik, Psychologie, Ethik und Religionslehre
WA	Wissenschaftliches Arbeiten
WR	Fachbereich Wirtschaft und Recht
Z	Zusätzliches

Kompetenzbereiche Erstsprache

ALLG_LV	Allgemeines Leseverstehen (zur Orientierung lesen)
DETAIL_LV	Detailliertes Leseverstehen (Information & Argumentation verstehen)
GRAM	Sprachreflexion/Grammatische Kompetenz und Orthografie

TOT_SPRACHE	Erstsprache gesamt (ohne Grammatik)
WORTS	Sprachreflexion/Wortschatz

Kompetenzbereiche Mathematik

AF	Analysis/Funktionen
AI	Analysis/Infinitesimalrechnung
AR	Analysis/Folgen und Reihen
GG	Geometrie/elementare und analytische Geometrie
GT	Geometrie/Trigonometrie
GV	Geometrie/Vektorrechnung und Vektorgeometrie
SK	Stochastik/Kombinatorik
SS	Stochastik/Statistik
SW	Stochastik/Wahrscheinlichkeit
XA	Vermischtes/elementare Algebra
XL	Vermischtes/Logik
XM	Vermischtes/Mengenlehre
ANALYSIS	Analytisches (Funktionen, Infinitesimalrechnung, Folgen und Reihen)
GEOMETRIE	Geometrisches (Trigonometrie, analytische Geometrie, Vektorrechnung)
M_VERSCH	Diverses/Vermischtes (Mengenlehre/diskrete Mathematik, Logik, elementare Algebra)
STOCHASTIK	Stochastisches (Wahrscheinlichkeitsrechnung, Kombinatorik, Statistik)
TOT_MATH	Mathematik gesamt

Kompetenzbereiche Biologie

B_EVO	Evolution (Entwicklungsprozesse – Evolution und Zukunftsfragen)
B_GEN	Genetik (Grundlagen der molekularen Genetik, Anwendung moderner biologischer Erkenntnisse und Methoden)
B_INFO	Informationsverarbeitung (Kommunikation zwischen Zellen, neuronale Informationsverarbeitung, Wahrnehmung)
B_OEKO	Ökologie (Vernetzte Systeme – Ökologie und Nachhaltigkeit)
B_STOFF	Stoffwechsel (Stoff- und Energiewechsel: Notwendigkeit und Wege der Energieumwandlung, Assimilation und Dissimilation im Zusammenhang von zellulären Strukturen und Organismus; Molekulare Steuerung von Stoffwechsel: Enzymatik)
B_STRU	Struktur und Funktion (Bau und Funktion von Zellen, Geweben und Organen; funktionsbezogene Differenzierungen)
TOT_BIO	Biologietest gesamt

Kompetenzbereiche Überfachlicher Fähigkeitstest

U_DIAGR	Diagramme und Tabellen
U_GESAMT	Totalergebnisse des überfachlichen Fähigkeitstests

U_NAT	Naturwissenschaftliches Grundverständnis
U_QUANT	Quantitative und formale Probleme
U_TEXT	Textverständnis

Parameter Zusatzbefragung

Pruefvo	Prüfungsvorbereitung
Zeiteff	Zeiteffizienz
Zeitpla	Zeitplanung

Notenwerte Erstprache

E_ErfNote	Erstsprache Erfahrungsnote
E_MatNote	Gesamtnote der Maturitätsprüfung in Erstsprache
E_müPr	Note der mündlichen Maturitätsprüfung in Erstsprache
E_schrPr	Note der schriftlichen Maturitätsprüfung in Erstsprache

Notenwerte Mathematik

Ma_ErfNote	Mathematik Erfahrungsnote
Ma_MatNote	Gesamtnote der Maturitätsprüfung in Mathematik
Ma_müPr	Note der mündlichen Maturitätsprüfung in Mathematik
Ma_schrPr	Note der schriftlichen Maturitätsprüfung in Mathematik
MEF_schrPr	Note der schriftlichen Maturitätsprüfung im Ergänzungsfach Mathematik
MEF_müPr	Note der mündlichen Maturitätsprüfung im Ergänzungsfach Mathematik naturwissenschaftlichen Unterrichts e.V.
MPhyS_MatPr	Note der Maturitätsprüfung im Schwerpunktfach Physik und Anwendungen der Mathematik
MphyS_müPr	Note der mündlichen Prüfung des Schwerpunktfaches Physik und Anwendungen der Mathematik
MphyS_schrPr	Note der schriftlichen Prüfung des Schwerpunktfaches Physik und Anwendungen der Mathematik
Ph_ErfNote	Physik Erfahrungsnote

Notenwerte Biologie

Bi_ChS_schrPr	Note der schriftlichen Maturitätsprüfung im Schwerpunktfach Biologie/Chemie
Bi_ErfNote	Biologie Erfahrungsnote
Bi_müPR	Note der mündlichen Maturitätsprüfung in Biologie
Bi_schrPr	Note der schriftlichen Maturitätsprüfung in Biologie
BiChS_müPr	Note der mündlichen Maturitätsprüfung im Schwerpunktfach Biologie/Chemie
BiEF_müPr	Note der mündlichen Maturitätsprüfung im Ergänzungsfach Biologie
BiEF_schrPr	Note der schriftlichen Maturitätsprüfung im Ergänzungsfach Biologie

Ch_ErfNote	Chemie Erfahrungsnote
NW_MatNote	Gesamtnote der Maturität im Fachbereich Naturwissenschaften

Schweizer Kantone

AG	Kanton Aargau
AI	Kanton Appenzell Innerrhoden
AR	Kanton Appenzell Ausserrhoden
BEd	Kanton Bern deutschsprachig
BEf	Kanton Bern französischsprachig
BS	Kanton Basel Stadt
FR	Kanton Fribourg
GL	Kanton Glarus
GR	Kanton Graubünden
JU	Kanton Jura
LU	Kanton Luzern
NE	Kanton Neuenburg
NW	Kanton Nidwalden
OW	Kanton Obwalden
SG	Kanton St. Gallen
SH	Kanton Schaffhausen
SO	Kanton Solothurn
SZ	Kanton Schwyz
TG	Kanton Thurgau
UR	Kanton Uri
VD	Kanton Waadt
VS	Kanton Wallis
VSd	Kanton Wallis deutschsprachig
VSf	Kanton Wallis französischsprachig
ZG	Kanton Zug

Allgemeine Abkürzungen

AS-Sequenz	Aminosäurensequenz
ATP	Adenosintriphosphat
BMBF	Bundesministerium für Bildung und Forschung, Bundesrepublik Deutschland
CRUS	Rektorenkonferenz der Schweizer Universitäten
DESI	Deutsch-Englisch-Schülerleistungen-International
DGPuK	Deutsche Gesellschaft für Publizistik- und Kommunikationswissenschaft
DNS	Desoxyribonukleinsäure
EDK	Schweizerische Konferenz der kantonalen Erziehungsdirektoren
EMS	Eignungstest für das Medizinstudium

EPA	Einheitliche Prüfungsanforderungen in der Abiturprüfung Biologie
EPFL	Ecole Polytechnique Fédérale de Lausanne
ETH, ETHZ	Eidgenössische Technische Hochschule (Zürich)
EVAMAR I	Evaluation der Maturitätsreform 95, Phase I (2004)
EVAMAR II	Evaluation der Maturitätsreform 95, Phase II (2008)
FHNW	Fachhochschule Nordwestschweiz, Pädagogische Hochschule, Institut Forschung und Entwicklung
geg.	gegeben
GER	Gemeinsamer Europäischer Referenzrahmen für Sprachen
ges.	gesucht
gew.	gewichtet
GRE	Biology Graduate Record Examination
HarmoS	Interkantonale Vereinbarung über die Harmonisierung der obligatorischen Schule
HRM	human resources management (Personalmanagement)
i.d.F.	in dieser Form
i.d.R.	in der Regel
IBE	Institut für Bildungsevaluation, Universität Zürich
IGB	Institut für Gymnasial- und Berufspädagogik der Universität Zürich
IRDP	Institut de recherche et de documentation pédagogique, Neuchâtel
Kat.	Kategorie
KMK	Kultusministerkonferenz, Bundesrepublik Deutschland
KSGR	Konferenz Schweizerischer Gymnasialrektorinnen und Gymnasialrektoren
MA	Mittelalter
MAR 95	Maturitätsanerkennungsreglement von 1995
math.	mathematisch
MAV 68	Maturitätsanerkennungsverordnung von 1968
Max A	Erreichte Maximalpunktzahl Variante A
Max P	Höchste Personenfähigkeit Variante P
MCAT	Medical College Admission Test
Min P	Tiefste Personenfähigkeit Variante P
Mittel A	Mittelwerte nach Variante A
Mittel P	Mittelwerte nach Variante P
MNU	Deutscher Verein zur Förderung des mathematischen und naturwissenschaftlichen Unterrichts
NCTM	National Council of Teachers of Mathematics
NRW	Nordrhein-Westfalen
OECD	Organisation for Economic Co-operation and Development
OLAT	Online learning and training – Lernplattform
PISA	Programme for International Student Assessment
PSE	Periodensystem der Elemente
RLP	Rahmenlehrplan

RNA	Ribonucleinsäure
RVK	Regensburger Verbundklassifikation
SBF	Staatssekretariat für Bildung und Forschung
SD	Standardabweichung (standard deviation)
SekI, SI	Sekundarstufe I
SekII, SII	Sekundarstufe II
Sig.	Signifikanz
SPF	Schwerpunktfach
TIMSS	Third International Mathematics and Science Study
TMS	Test für medizinische Studiengänge (Deutschland)
TP	Teilprojekt
UFT	überfachlicher Fähigkeitstest
USR	Ufficio studi e ricerche, Bellinzona
UZH	Universität Zürich
Variante A	Normierung auf max. 20 Punkte pro Bereich
Variante P	Wert null entspricht einer „mittleren“ Fähigkeit
vs.	versus (im Vergleich zu)
VSEPR	Valenzschalen-Elektronenpaarabstossungs-Modell (Valence shell electron pair repulsion)
VSG	Verein Schweizerischer Gymnasiallehrerinnen und Gymnasiallehrer
ZTD	Zentrum für Testentwicklung und Diagnostik, Fribourg