

Studienberatung im deutschen
Hochschulsystem auf der Basis
psychologischer Tests:
Studienfachprofile, Vorhersagevalidität
und Akzeptanz

Von der Philosophischen Fakultät der
Rheinisch-Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades eines Doktors der
Philosophie genehmigte Dissertation

vorgelegt von
Dipl.-Psych. Alexander Zimmerhofer aus Duisburg

Berichter:
Univ.-Prof. Dr. Lutz F. Hornke
Univ.-Prof. Dr. Klaus Willmes-v. Hinckeldey

Tag der mündlichen Prüfung:
23. Mai 2008

Zu Dank verpflichtet bin ich allen, die mich auf dem Weg zu dieser Arbeit unterstützt haben:

Mein Betreuer Prof. Dr. Lutz F. Hornke bereitete weitsichtig als Initiator den Boden für diese Arbeit, aber auch für weitere Projekte zu webbasierten Studienberatungssystemen. Mit einem stets offenen Ohr betreute er nicht nur die vorliegende wissenschaftliche Arbeit, sondern half auch andersartige Schwierigkeiten zu umschiffen. Viola Weber unterstützte im Rahmen ihrer Diplomarbeit tatkräftig die Entwicklung der ersten Version des Self-Assessments. Prof. Dr. Ulrik Schroeder und Eva Giani ermöglichten durch ihren Einsatz die Kriteriumsvalidierung des Self-Assessments.

Meiner Ehefrau und meinen Eltern zolle ich Dank für ihre bedingungslose Unterstützung, auch in schwereren Zeiten. Darüber hinaus bin ich meinen ehemaligen Kolleginnen und Kollegen am Lehrstuhl für Betriebs- und Organisationspsychologie der RWTH Aachen zu Dank verpflichtet. Insbesondere möchte ich Dr. Jonas Lang erwähnen, der mit zahlreichen erquickenden Diskussionen half. Verena Heukamp und Annika Milbradt halfen mit vielen Anregungen akribisch, die Texte verständlicher und fehlerfreier drucken zu können.

Abschließend geht ein außerordentlicher Dank an die mehr als 10 000 Personen, die die empirische Überprüfung durch ihre Teilnahme am Self-Assessment erst ermöglicht haben und die durch ihre bloße Teilnahme, aber auch durch die zahlreichen Rückmeldungen gezeigt haben, dass ein wichtiges Instrument geschaffen wurde.

Zusammenfassung

Mit dem Ziel, langfristig die Studienerfolgsquoten insbesondere in den Studienfächern Informatik und Elektrotechnik zu steigern, wurde an der RWTH Aachen ab dem Jahr 2002 ein webbasiertes Studienberatungssystem (ein so genanntes Self-Assessment) auf der Grundlage von kognitiven Tests und nicht-kognitiven Fragebogen entwickelt. In dieser Arbeit werden zuerst der Hintergrund der Entwicklung, der Aufbau der Testbatterie und die technische Umsetzung dargelegt, aber auch empirische Ergebnisse zur Testgüte und zur Akzeptanz bei den Teilnehmenden berichtet. In einem Extremgruppenvergleich von Informatik-Interessierten und geisteswissenschaftlich Interessierten wird untersucht, welche Bestandteile der Testbatterie in der Lage sind, zwischen diesen Gruppen zu differenzieren. Dabei können die nicht-kognitiven Fragebogen mehr Varianz in der Zuordnung aufklären, als dies die kognitiven Tests ermöglichen. Bei einer Vollerhebung unter Studienanfängerinnen und -anfängern der Informatik an der RWTH Aachen ($n = 594$) wird zudem die Validität der Testbatterie anhand zweier Kriterien (Studiennoten und Anzahl bestandener Prüfungen) untersucht. Sowohl die kognitiven Test-Bestandteile ($\Delta R = 0,196$) alleine als auch eine Kombination aus kognitiven und nicht-kognitiven Bestandteilen ($\Delta R = 0,291$) zeigen eine inkrementelle Validität für die Vorhersage der späteren Studiennoten über die Abiturnote ($R = 0,304$) hinaus. Auf der berichteten empirischen Grundlage werden Ratschläge zur Entwicklung und Umsetzung webbasierter Studienberatungssysteme, aber auch von Studieneignungstests für die Studierendenauswahl gegeben.

Schlagwörter: Studieneignung, Studienneigung, Studienerfolgsprognose, webbasierte Studienberatung, Self-Assessment

Abstract

The RWTH Aachen University has been developing a web-based self-assessment student guidance system since 2002. The programme is based on cognitive tests and non-cognitive questionnaires and has the long-term goal of increasing success rates, in particular in Computer Sciences and Electrical Engineering. This work provides information about the background to the development, the construction of the test battery and technical implementation as well as empirical evidence regarding the quality of the test and acceptance by the participants. An extreme group comparison between interested Computer Sciences students and Humanities students was carried out to determine which components of the test battery were capable of differentiating between these groups. The non-cognitive questionnaires were then able to explain more variance in the correlation than was possible with the cognitive tests.

In addition, the validity of the test battery was also examined by means of a census of new Computer Sciences students at the RWTH Aachen University ($n = 594$) using two criteria: student grades and the number of exams passed. The cognitive test components ($\Delta R = 0,196$) alone and a combination of the cognitive and non-cognitive components ($\Delta R = 0,291$) demonstrated an incremental validity for predicting university grades from the Abitur (German high school graduation) grades ($R = 0,304$).

The reported empirical basis will be used to form recommendations for the development and implementation of web-based student guidance systems as well as scholastic aptitude tests for student selection.

Keywords: Scholastic aptitude, prediction of university success, web-based student guidance, self-assessment

Inhaltsverzeichnis

Tabellenverzeichnis	XIII
Abbildungsverzeichnis	XV
Abkürzungsverzeichnis	XVII
1 Das deutsche Hochschulsystem in der Debatte	1
1.1 Einführung	1
1.2 Qualitätsproblematik im deutschen Bildungssystem	2
1.3 Quantitätsproblematik im deutschen Bildungssystem	3
1.4 Studienabbruch und Studienfachwechsel	6
1.4.1 Studienabbruch – Definitionen und Ausmaße	6
1.4.2 Schwundquote – Definitionen und Ausmaße	9
1.4.3 Internationaler Vergleich der Studienabbruchquoten	9
1.4.4 Auswirkungen von Studienabbruch	11
1.4.5 Gründe für Studienabbruch	13
1.5 Zusammenfassung und Ausblick	20
2 Eignungsbeurteilungen in der Fremd- und Selbstselektion von Studieninteressierten	21
2.1 Die irrationalen Entscheidungen bei einer Studienfachwahl	21
2.2 Die rationale Studienfachwahl auf Basis der Eignung und Neigung	24
2.2.1 Die Person-Studienfach-Passung	24
2.2.2 Die allgemeine und fachspezifische Studierfähigkeit	26
2.3 Fremdselektion durch die Hochschulen	30
2.4 Selbstselektion durch den Bewerber	31
2.5 Internetgestützte Berufs- und Studienberatung	33
2.6 Zusammenfassung	39
3 Kriterien, Verfahren und Prognostizierbarkeit des Studienerfolgs	41
3.1 Studienerfolg - Definition möglicher Kriterien	41
3.1.1 Studienabschluss versus Studienabbruch	42
3.1.2 Studiennoten	43

3.1.3	Studiendauer	44
3.1.4	Studienzufriedenheit	45
3.1.5	Zusammenfassung	46
3.2	Hochschulreifezeugnis als Prädiktor	47
3.3	Kognitive Verfahren als Prädiktoren	52
3.3.1	Allgemeine Intelligenz	52
3.3.2	Spezifische Komponenten der Intelligenz	53
3.4	Nicht-kognitive Verfahren als Prädiktoren	57
3.4.1	Big-Five-Persönlichkeitsdimensionen	58
3.4.2	Rubikon-Modell der Handlungsphasen	60
3.4.3	Berufliche Interessen	63
3.4.4	Extrinsische Lernmotivation	71
3.4.5	Selbstwirksamkeit	73
3.4.6	Handlungs- versus Lageorientierung	75
3.5	Zusammenfassung	77
4	Ziele der vorliegenden Arbeit	79
5	Studie 1: Entwicklung und Evaluation einer Testbatterie zur Studienberatung	83
5.1	Einleitung	83
5.2	Fragestellung	84
5.3	Methoden	84
5.3.1	Stichprobe	84
5.3.2	Verwandte Tests	87
5.3.3	Technische Realisierung	97
5.3.4	Verwandte Auswertungsverfahren	98
5.4	Ergebnisse	99
5.4.1	Rohwerteverteilung der Skalen	99
5.4.2	Interne Konsistenz der Skalen	102
5.4.3	Interkorrelationen der Subskalen	103
5.4.4	Reliabilität der Self-Assessment-Batterie und der Teilnehmerprofile	106
5.4.5	Faktorielle Struktur der Self-Assessment-Batterie	106
5.5	Diskussion	108
6	Studie 2: Profile unterschiedlicher Studienfächer	111
6.1	Einleitung	111
6.2	Fragestellung	112
6.3	Methoden	114
6.3.1	Stichprobe	114
6.3.2	Verwandte Testverfahren	114

6.3.3	Technische Realisierung	115
6.3.4	Verwandte Auswertungsverfahren	115
6.4	Ergebnisse	117
6.5	Diskussion	126
7	Studie 3: Kriteriumsvalidierung der Testbatterie	129
7.1	Einleitung	129
7.2	Fragestellung	130
7.3	Methoden	131
7.3.1	Stichprobe	131
7.3.2	Verwandte Testverfahren	132
7.3.3	Technische Realisierung	132
7.3.4	Verwandte Auswertungsverfahren	133
7.4	Ergebnisse	135
7.4.1	Korrelationen der Kriterien mit den Prädiktoren	135
7.4.2	Vorhersage der Studiennoten im Hauptfach Informatik	138
7.4.3	Vorhersage der Anzahl bestandener Prüfungen im Hauptfach Informatik	144
7.4.4	Vorhersage der Studiennoten im Nebenfach Informatik	147
7.4.5	Vorhersage der Anzahl bestandener Prüfungen im Nebenfach Informatik	147
7.5	Diskussion	148
8	Studie 4: Akzeptanz	155
8.1	Einleitung	155
8.2	Methoden	156
8.2.1	Stichprobe	156
8.2.2	Verwandte Testverfahren	156
8.2.3	Kontrolle des selektiven Dropouts	158
8.2.4	Technische Realisierung	158
8.3	Ergebnisse	160
8.3.1	Zugriffs- und Teilnehmerzahlen	160
8.3.2	Akzeptanzbefragung	164
8.4	Diskussion	167
9	Diskussion	171
9.1	Self-Assessments als Instrument zur Verbesserung der Studienberatung	171
9.2	Eignungsdiagnostische Fragestellungen	175
10	Literaturverzeichnis	179
11	Anhang	213

Tabellenverzeichnis

3.1	Beschreibung der Interessen-Dimensionen nach Holland	69
5.1	Itemanzahl, Mittelwerte, Standardabweichung und interne Konsistenz der konstruierten Skalen	101
5.2	Interkorrelationen der Skalen im Self-Assessment	104
5.3	Interkorrelationen der Skalen im Self-Assessment, doppelt minderungskorrigiert	105
5.4	Komponentenmatrix der Testbatterie nach Varimax-Rotation	106
6.1	Vergleich der Testergebnisse von Studieninteressierten der Geisteswissenschaften und der Informatik, Elektrotechnik und Technische Informatik	118
6.2	Vergleich der Testergebnisse von Studieninteressierten der Elektrotechnik und der Informatik	118
6.3	Vergleich der Testergebnisse von Studieninteressierten beider Geschlechter	120
6.4	Logistische Regression zur Vorhersage des Studieninteresses	124
7.1	Statistischer Zusammenhang (Pearson-Korrelation) zwischen den Prädiktoren (Zeilen) und den Kriterien des Studienerfolgs (Spalten)	137
7.2	Vorhersage der mittleren Vordiplomnote im Hauptfach, getrennt nach kognitiven und nicht-kognitiven Bestandteilen	140
7.3	Vorhersage der mittleren Vordiplomnote im Hauptfach, durch Tests und Fragebogen getrennt	141
7.4	Standardisierte β -Gewichte zur Vorhersage der Studiennoten (Bootstrapping)	141
7.5	Vorhersage der mittleren Vordiplomnote im Hauptfach über das Abitur hinaus	142
7.6	Multiple Regression zur Vorhersage der Anzahl bestandener Vordiplomprüfungen im Hauptfach Informatik	145
7.7	Standardisierte β -Gewichte zur Vorhersage der Prüfungsanzahl	146
7.8	Multiple Regression zur Vorhersage der Anzahl bestandener Vordiplomprüfungen im Hauptfach Informatik	147
7.9	Mittlere multiple Korrelation zur Vorhersage der Vordiplomnoten und Anzahl bestandener Prüfungen im Hauptfach und Nebenfach Informatik	148
8.1	Mittlere Itembeantwortungen und die interne Konsistenz der fünf Akzeptanzskalen	165
8.2	Interkorrelationen zwischen Akzeptanzscores und Testscores	166

Abbildungsverzeichnis

1.1	Studienabbruchmotive und deren Bedingungsfaktoren	15
2.1	Person-Studienfach-Passung als Voraussetzung für den Studienerfolg	25
2.2	Darstellung allgemeiner und fachspezifischer Studierfähigkeit	29
3.1	Das Rubikonmodell des Handelns	62
3.2	Hexagon der beruflichen Interessen	67
3.3	Wirksamkeitserwartung und Ergebniserwartung nach Bandura	73
5.1	Screepplot der Originaldaten und der simulierten Daten im Rahmen der Parallel- analyse	108
5.2	Ergebnis der exploratorischen Faktorenanalyse der gesamten Testbatterie	108
6.1	Geschlechtsunterschiede SW	122
6.2	Geschlechtsunterschiede ELM	122
6.3	Geschlechtsunterschiede HK	123
6.4	Geschlechtsunterschiede INT	123
6.5	Geschlechtsunterschiede MAT	123
6.6	Geschlechtsunterschiede LOG	123
6.7	Geschlechtsunterschiede MATH	123
6.8	Sensitivität und Komplement der Spezifität zur Unterscheidung von Interessen- tengruppen	124
6.9	Sensitivität und Komplement der Spezifität zur Unterscheidung von Interessen- tengruppen	126
7.1	Prüfung der Residuen auf Normalverteilung	138
7.2	Prüfung der Residuen auf Heteroskedastizität	138
7.3	Verteilung der multiplen Korrelation zwischen der Testbatterie und der mittleren Studiennote	141
7.4	Sensitivität und Komplement der Spezifität der Prädiktoren für die Vorhersage der späteren Studiennoten	144
7.5	Verteilung der multiplen Korrelation zwischen den Prädiktoren und der Anzahl der bestandenen Prüfungen	146

8.1	Besuche des Self-Assessments in den Jahren 2002 bis 2006	160
8.2	Anzahl der Seitenabfragen (Requests) im Tagesverlauf	162
8.3	Überblick über die Anzahl der Besuche, Testungen und Testteilnehmer	163
11.1	Scoreverteilung Skala INT-R	214
11.2	Q-Q-Diagramm Skala INT-R	214
11.3	Scoreverteilung Skala INT-I	214
11.4	Q-Q-Diagramm Skala INT-I	214
11.5	Scoreverteilung Skala INT-A	215
11.6	Q-Q-Diagramm Skala INT-A	215
11.7	Scoreverteilung Skala INT-S	215
11.8	Q-Q-Diagramm Skala INT-S	215
11.9	Scoreverteilung Skala INT-E	216
11.10	Q-Q-Diagramm Skala INT-E	216
11.11	Scoreverteilung Skala INT-C	216
11.12	Q-Q-Diagramm Skala INT-C	216
11.13	Scoreverteilung Skala SW	219
11.14	Q-Q-Diagramm Skala SW	219
11.15	Scoreverteilung Skala ELM	220
11.16	Q-Q-Diagramm Skala ELM	220
11.17	Scoreverteilung Skala HK	221
11.18	Q-Q-Diagramm Skala HK	221
11.19	Scoreverteilung Skala HOM	221
11.20	Q-Q-Diagramm Skala HOM	221
11.21	Scoreverteilung Skala HOP	222
11.22	Q-Q-Diagramm Skala HOP	222
11.23	Scoreverteilung Skala HOT	222
11.24	Q-Q-Diagramm Skala HOT	222
11.25	Scoreverteilung und Q-Q-Diagramm Skala LOG	226
11.36	Scoreverteilung und Q-Q-Diagramm Skala MAT	232
11.47	Scoreverteilung und Q-Q-Diagramm Skala MATH	238

Abkürzungsverzeichnis

ACT.....	ehemals The American College Testing Program, Non-Profit-Organisation, entwickelt u.a. Studierfähigkeitstests, die von vielen US-Colleges als Eingangsprüfung verwandt werden
BAFöG	Bundesausbildungsförderungsgesetz
BIS	Berliner-Intelligenz-Strukturmodell
CBTI.....	Computer-based test interpretation
df.....	Anzahl der Freiheitsgrade
ELM.....	Extrinsische Lernmotivation
GMA	Generelle mentale Fähigkeit
GPA	Grade Point Average, US-amerikanischer Zeugnisnotendurchschnitt
HIS	Hochschul-Informationssystem GmbH
HOM	Handlungskontrolle – Handlungsorientierung nach Misserfolg
HOP.....	Handlungskontrolle – Handlungsorientierung bei der Handlungsplanung
HOT.....	Handlungskontrolle – Handlungsorientierung bei der Tätigkeitsausführung
INT.....	Interessenfragebogen
INT-A	Interessenskala artistic – künstlerisch
INT-C	Interessenskala conventional – verwaltend
INT-E	Interessenskala enterprising – unternehmerisch
INT-I.....	Interessenskala investigative – wissenschaftlich-analytisch
INT-R	Interessenskala realistic – praktisch-technisch
INT-S.....	Interessenskala social – erzieherisch-pflegend
KI	Konfidenzintervall
LOG.....	Testverfahren zur Erfassung schlussfolgernden Denkens, Logiktest
M.....	Arithmetisches Mittel

MAT.....	Matrizentest
MATH.....	Testverfahren zur Erfassung mathematischer Fähigkeiten, Mathe- test
MD.....	Median
OECD.....	Organisation für Wirtschaftliche Zusammenarbeit und Entwicklung
P-E-Fit.....	Personen-Umwelt-Fit
PISA.....	Programme for International Student Assessment
SAT.....	Allgemeiner Studierfähigkeitstest für die Zulassung insbesondere an US-amerikanischen Colleges, ehemals Scholastic Aptitude Test
SD.....	Standardabweichung
SW.....	Selbstwirksamkeit
Tertiärbereich A.....	Bereich des universitären Bildungssystems, d.h. Hochschule bzw. Fachhochschule
TIMSS.....	Third International Mathematics and Science Study
ZVS.....	Zentralstelle für die Vergabe von Studienplätzen

Kapitel 1

Das deutsche Hochschulsystem in der Debatte

1.1 Einführung

Angesichts der zunehmenden Wettbewerbsfähigkeit anderer Volkswirtschaften und attraktiver Produktionsstandorte mit geringeren Fertigungsgemeinkosten in der unmittelbaren Nachbarschaft werden westliche Industrieländer – und somit auch Deutschland – ihre internationale Wettbewerbsposition vor allem anderen mit intelligenten, hochwertigen und innovativen Gütern, Dienstleistungen und Produktionsverfahren sichern müssen (Buttler & Tessaring, 1993). Die Produktion günstiger Massenprodukte wurde bereits beziehungsweise wird in den nächsten Jahren vermehrt in Niedriglohnländer verlagert. Zu groß sind die Unterschiede der Personalkosten, zu gering sind die erzielbaren Qualitäts- beziehungsweise Effizienzunterschiede bei diesen Produkten. Eine besondere Bedeutung für den volkswirtschaftlichen Erfolg gerade im Segment der hochpreisigen, innovativen Produkte spielt die Quantität und Qualität von Fachkräften (Lutz & Meil, 2000), die derzeit in der benötigten Anzahl in den Niedriglohnländern nicht mit vergleichbaren Fähigkeiten, Fertigkeiten und Kenntnissen zur Verfügung stehen. Eine entscheidende Ressource ist daher das so genannte *individuelle Humankapital* oder *Humanvermögen*. Dieses stellt den Bestand „an Fähigkeiten und Kenntnissen dar, die der Einzelne besitzt oder – normalerweise durch Bildung oder Ausbildung – (weiter-) entwickelt und sodann als Gegenleistung für ein Einkommen auf dem Arbeitsmarkt anbietet“ (Organisation für Wirtschaftliche Zusammenarbeit und Entwicklung, 2004, S. 30). Davon unterschieden wird das gesamtgesellschaftliche Humankapital, das der Summe des produktiven Bildungsstandes eines gesamten Landes in Form von Arbeitskraft, Fähigkeiten, Fertigkeiten und Wissen gleicht.

Ein Ausbau dieser volkswirtschaftlichen Ressource, insbesondere durch eine Steigerung der Qualität und Quantität der inländischen Fachkräfte über eine Verbesserung an Schulen und

Hochschulen, ist eine Voraussetzung, um den Standort Deutschland international wettbewerbsfähig zu halten (Braun, 2001). Investitionen in das Bildungssystem sichern wiederum Investitionen in Sachkapital, fördern die Entwicklung und Verbreitung neuer Technologien, erhöhen die Arbeitsproduktivität, verbessern den Gesundheitszustand der Bevölkerung, senken die Kriminalitätsrate, stärken den sozialen Zusammenhalt, steigern das politische Engagement, wirken sich somit positiv auf das Wirtschaftswachstum des Landes aus und senken die Arbeitslosigkeit (Fuchs, Schnur & Zika, 2005; Organisation für Wirtschaftliche Zusammenarbeit und Entwicklung, 2004; Reinberg & Hummel, 2003a; Reinberg & Hummel, 2005; Reinberg & Schreyer, 2003). Dies macht deutlich, dass Bildung eben nicht nur ein Konsumgut, sondern eine Investition für eine Nation (Becker, 1964) und für das Individuum darstellt (Organisation für Wirtschaftliche Zusammenarbeit und Entwicklung, 2004), wobei die Rentabilität von Humankapitalinvestitionen gar der von Sachinvestitionen entspricht (Fuente & Ciccone, 2003).

Stehen dem Arbeitsmarkt in Deutschland nicht genügend hochqualifizierte Arbeitskräfte zur Verfügung, zeigen sich gegenläufige Effekte: So kann aus dem knappen Faktor Bildung beziehungsweise technologischem Wissen rasch eine Ursache für ein Innovationshemmnis werden (Bundesministerium für Bildung und Forschung, 2004b; Licht, 2002; Reinberg & Hummel, 2003b). Gerade aus diesem Grund besteht die Aufgabe des deutschen Bildungssystems in der effizienten Ausbildung auf einem hohen internationalen Niveau. Aber sowohl über die Qualität (siehe Abschnitt 1.2) von Schülern¹, Studenten und Absolventen als auch über die Quantität (siehe Abschnitt 1.3) von Akademikern wird in Deutschland heftig diskutiert. Die folgenden beiden Abschnitte werden diese Kritik näher beleuchten.

1.2 Qualitätsproblematik im deutschen Bildungssystem

Industrieunternehmen in Deutschland klagen verstärkt über die mangelhafte schulische Vorbereitung von Auszubildenden beziehungsweise Bewerbern (Bundesvereinigung der Deutschen Arbeitgeberverbände, 2003; Zweites Deutsches Fernsehen, 2005), aber auch über mangelnde Kompetenzen von Universitätsabsolventen (Wissenschaftsrat, 1999). Auch Hochschuldozenten sprechen vielen Studenten die Eignung, aber auch das nötige Vorwissen für einen erfolgreichen Universitätsbesuch ab (Konegen-Grenier, 2001; Spiegel-Online, 2002). So ist nur etwa jeder zehnte Befragte ($n = 1435$) unter Universitätsprofessoren der Meinung, dass Schulabsolventen mit Abitur generell reif für die Aufnahme eines Studiums sind (Konegen-Grenier, 2001). Indirekt lässt sich ebenso ableiten, dass nur wenige der Befragten das Ziel des Zeugnisses, die allgemeine Hochschulreife zu bescheinigen, erreicht sehen. Auch Umfragen unter Studenten zeigen durchaus vergleichbare Ergebnisse: Etwa ein Drittel der befragten Erstsemester ($n = ca.$

¹ Im Folgenden wird zur besseren Lesbarkeit stets die männliche Form von Personenbezeichnungen verwandt. Sie steht stellvertretend für Personen beider Geschlechter.

9 000) waren der Meinung, dass sie seitens der Schulen unzureichend vorbereitet wurden (Lewin, Heublein & Sommer, 2000), also die Hochschulreife nicht besaßen. Insbesondere im Bereich des mathematischen Vorwissens (59 % der Befragten) wurden Lücken zugegeben. Nur jeder dritte Anfänger fühlte sich durch die Schule gut oder gar sehr gut auf ein universitäres Studium vorbereitet.

Diese eher subjektiven Einschätzungen seitens der Hochschulprofessoren und ihrer Studierenden wurden gerade nach den Bildungsstudien der Organisation for Economic Co-operation and Development (OECD) von den Medien interessiert aufgenommen (Spiegel-Online, 2002). Seit der Veröffentlichung der TIMS- (Third International Mathematics and Science Study) und der ersten PISA-Studie (Programme for International Student Assessment) im Jahre 2001 ist die Bildungspolitik wieder verstärkt in den Fokus der Öffentlichkeit geraten. Medien titelten als Reaktion auf die gerade einmal mittelmäßigen Resultate deutscher Schüler aufgebracht: „Das Land der Dichter und Denker – abgehängt“ (Spiegel-Online, 2001), „Deutschland wieder unter Mittelmaß“ (Deutsche Presseagentur, 2004), „Die Bilanz des Schreckens“ (Die Zeit, 2003) oder „Die Schule brennt“ (Die Zeit, 2001). Die internationale Vergleichsstudie (Artelt et al., 2001) bescheinigte, dass deutsche Schülerinnen und Schüler eine im OECD-Vergleich unterdurchschnittliche Lesekompetenz ($M = 484$, $SD = 111$ bei PISA 2000; OECD-Mittel stets $M = 500$, $SD = 100$) besaßen. Der Abstand zur internationalen Spitzengruppe (Finnland, Kanada) betrug dabei etwa eine halbe Standardabweichung. Auch bei den Mathematikleistungen ($M = 490$, $SD = 103$) und bei der naturwissenschaftlichen Grundbildung ($M = 487$, $SD = 102$) befand sich Deutschland im unteren Mittelfeld. Bei allen Kompetenzarten wurden während des zweiten Erhebungszeitpunktes im Jahr 2003 (Prenzel et al., 2004) zwar leichte, oftmals allerdings nicht signifikante Zunahmen verzeichnet.

Die nachweislich im Mittel mäßige Qualität von Schülern, aber auch von Abiturienten stellen ein eklatantes Problem für den Wissenschafts- und Forschungsstandort Deutschland dar. Neue Ansätze werden in Deutschland gesucht werden müssen, die Qualität der Ausbildung in den Schulen zu verbessern und so die Diskrepanz zu anderen Industrienationen zu vermindern. Genauso müssen neuartige Instrumente entwickelt und eingesetzt werden, die das bereits vorhandene individuelle Humankapital zutreffend erkennen, aber auch zweckmäßig fördern. Das Abitur scheint zumindest alleine in vielen Fällen kein hinreichendes Kriterium für das Bildungspotenzial von Heranwachsenden zu sein (siehe weitere Ausführungen in Kapitel 3.2).

1.3 Quantitätsproblematik im deutschen Bildungssystem

Nicht nur die Qualität, sondern auch die Anzahl gut ausgebildeter Personen ist entscheidend für die Entwicklung einer Volkswirtschaft. Derzeit zeigt sich, dass der Neuaufbau beziehungsweise eine Verlegung sowohl von hochtechnisierten Produktions- als auch gerade von Forschungs-

und Entwicklungsabteilungen innerhalb oder nach Deutschland zunehmend durch die Anzahl an Fachkräften begrenzt ist. Die Industrieverbände – unter anderem der Bundesverband Digitale Wirtschaft (2006) oder der Bundesverband Informationswirtschaft, Telekommunikation und neue Medien (2006) –, aber auch das Institut der deutschen Wirtschaft (2005a; 2005b), das Institut für Arbeitsmarkt und Berufsforschung (2003b) und die Initiative D21 (Lückefett & Thomas, 2001) befürchten schon seit Jahren direkte Auswirkungen auf die Industrie.

Einer der wichtigsten Gründe für den heutigen Fachkräftemangel sind die geringen Geburtenraten in den letzten Jahrzehnten. Individualisierungstrends in der Gesellschaft wie die eigene Erwerbstätigkeit, die Selbstverwirklichung, aber auch das Streben nach Anerkennung außerhalb der Familie wurden in den letzten Jahren bedeutsamer (Bundesinstitut für Bevölkerungsforschung, 2004). Die individuellen Biographien sind heute – gerade bei Akademikern – nicht mehr eng mit der traditionellen sozialen Institution der *Familie* verbunden (Schmitt & Winkelmann, 2005). Dieser Wandel, aber auch die verbesserten Verhütungsmethoden, verdeutlichten sich in der so genannten „Europe’s Second Demographic Transition“, die von etwa 1965 bis Mitte der 1970er Jahre in Deutschland mit einem starken Geburtenrückgang einher ging und bis heute anhält: Auf dem derzeitig stabil niedrigen Geburtenniveau werden im Durchschnitt je Frau 1,4 Kinder geboren (Bundesinstitut für Bevölkerungsforschung, 2004). Somit lässt sich die Elterngeneration lediglich zu etwa 67 % ersetzen.

Eine überaus einschneidende Begebenheit, die zahlreiche Diskussionen über Geburtenrückgang, Verschwendung von Humanressourcen und Fachkräftemangel etc. einleitete, war die Veröffentlichung der Publikation „Die deutsche Bildungskatastrophe“ von Georg Picht (1965) (Kehm, 2004; Teichler, 2005). Diese Veröffentlichung regte nachhaltig eine Diskussion an, die die besondere Bedeutung des Bildungssystems für Deutschland und die Versäumnisse in diesem Bereich aufdeckte. Picht sprach sich klar für eine Ausschöpfung der Begabungsreserven aus, um trotz eines Geburtenrückgangs das wirtschaftliche Wachstum Deutschlands zu sichern. Ansonsten droht, fasst man Pichts Aussagen zusammen, „die Bundesrepublik Deutschland wirtschaftlich auf den Stand eines Entwicklungslands zurückzufallen“ (Teichler, 2005, S. 9).

Angestoßen durch die zahlreichen Diskussionen war es seit den 1960er Jahren und ist es auch heute noch Ziel der Politik (Bundesministerium für Bildung und Forschung, 2006), mehr Heranwachsenden eine Studierberechtigung zu geben: 1950 erreichten lediglich 4,5 %, 1960 7,3 %, 1970 11 % und 1980 dann 18 % der Gleichaltrigen das Abitur (Wolter, 1989). Bis heute hat sich diese Tendenz fortgesetzt: Im Jahr 2003 haben 39,2 % der 18- bis 21-Jährigen die Hochschulbeziehungsweise die Fachhochschulreife erlangt. Innerhalb Deutschlands zeigt sich dabei jedoch eine beträchtliche Streuung dieser Studierberechtigtenquote² zwischen 28,0 % (Mecklenburg-Vorpommern) beziehungsweise 31,3 % (Bayern) und 44,0 % (Hessen) beziehungsweise 47,9 % (Nordrhein-Westfalen).

² Anteil der Studienberechtigten an der gesamten Bevölkerung von 18 bis unter 21 Jahren.

Mit dieser drastischen Steigerung der Schülerzahlen in den Gymnasien, aber auch der Studierendenzahlen, geht gleichzeitig ein Wandel der ehemaligen exklusiven Einrichtungen für Jahrgangsminderheiten, des so genannten „elite access system“, zu einem „mass access system“ einher (Wolter, 1989). Aus diesem Grund wird in vielen Debatten die Bildungsexpansion mit einer Entwertung beziehungsweise Inflationierung der Allgemeinen Hochschulreife verbunden (Schmidt-Klingenberg, 2002; Spiegel-Online, 2002; Spiewak, 2004). Grundidee dieser Kritik ist, dass sich die Zahl der Geeigneten nicht gleichermaßen wie die gestiegenen Abiturienten- beziehungsweise Studierendenzahlen verändert haben können. Somit lässt sich direkt daraus folgern, dass sich die Quote scheinbar nur durch ein Absenken der Qualität und somit auch durch eine Einschränkung der Auswahlfunktion an Schulen erreichen ließ. Durchaus fraglich ist allerdings, ob zum Beispiel bei historischen Quoten der Studierberechtigten zwischen 10 und 20 % bereits alle Geeigneten eine Allgemeine Hochschulreife erhalten haben. Erfahrungen in anderen Ländern zeigen, dass auch höhere Berechtigtenquoten ohne erkennbare Qualitätseinbußen erreicht werden können: Trotz der positiven Entwicklung der Studierberechtigtenquote lag Deutschland im Jahr 2002 immerhin etwa 20 Prozentpunkte unter dem OECD-Mittel von 54 % (Organisation für Wirtschaftliche Zusammenarbeit und Entwicklung, 2004).

Mehrfach wurde unter anderem von der OECD (zum Beispiel Organisation für Wirtschaftliche Zusammenarbeit und Entwicklung, 2005) oder ebenfalls von Picht (1965) auf ein weiteres Argument hingewiesen, das für eine mögliche Steigerung der Berechtigtenquote spricht: die soziale Selektion des deutschen Bildungssystems. Im Rahmen der PISA-Studien (Artelt et al., 2001) konnte nachgewiesen werden, dass Kindern sozial schwächerer Familien mit einer höheren Wahrscheinlichkeit die Hochschulreife verwehrt wird als Kindern aus bildungsnahen Schichten. Eine Problematik, die sich durch die Einführung von Studiengebühren noch weiter verstärken wird. Ein Kriterium für die Aufnahme eines Studiums sollte entsprechend alleine die Eignung für das Studienfach sein, nicht aber gesellschaftliche Rahmenbedingungen – wie eben der soziale Status der Eltern (operationalisierbar über den Schulabschluss, das Einkommen, die Berufe der Eltern etc.). Wird dieses Finanzierungsproblem z. B. durch verbesserte Stipendiensysteme gelöst, so steht eine Bildungsreserve aus den bisher unterrepräsentierten Bevölkerungsschichten zur Verfügung (Organisation für Wirtschaftliche Zusammenarbeit und Entwicklung, 2006), ganz ohne die Qualität der Ausbildung absenken zu müssen.

Ähnliches zeigt sich auch beim internationalen Vergleich der Hochschul-Abschlussquoten: Im OECD-Durchschnitt erreichen 31,8 % der Bevölkerung einen Abschluss des Tertiärbereichs A (Hochschul- beziehungsweise Fachhochschulabschluss) (Organisation für Wirtschaftliche Zusammenarbeit und Entwicklung, 2004). Spitzenwerte werden in Australien (45,4 %), Finnland (45,4 %), Polen (41,5 %) und Island (41,2 %) verzeichnet. In Deutschland hingegen schließen nur 19,2 % eines Jahrgangs ein Fachhochschul- oder Hochschulstudium ab³.

³ Gerade bei solchen Vergleichen müssen aber die Besonderheiten der jeweiligen Bildungssysteme berücksichtigt werden. So stehen in Deutschland Studieninteressierten mit der dualen Berufsausbildung attraktive

Es kann nicht angestrebtes Ziel in Deutschland sein, ohne Rücksicht auf die Qualität der Ausgebildeten die Akademikerquote zu steigern. Internationale Vergleiche lassen aber sehr wohl die Vermutung zu, dass eine Steigerung, die mit einer besseren Nutzung des derzeitigen Potenzials einhergeht, auch ohne Qualitätseinbußen möglich ist. Im Mittelpunkt muss dabei stehen, dass es – neben einer allgemeinen Diskussion der Zulassungsregelung zum Tertiärbereich A – auch immer wichtiger wird, geeignete Personen zu ermuntern, auch eine Hochschule zu besuchen, unabhängig vom sozialen Status der Eltern etc. Somit wird der Beratungsaspekt immer wichtiger, müssen doch auch die Jugendlichen selbst erst einmal die eigenen Qualitäten, das eigene, individuelle Humankapital erkennen. So sind neue, innovative Instrumente zu entwickeln, über Stärken und Schwächen, über mögliche Berufsfelder und deren Aussichten zu informieren und so vorhandene Begabungsreserven zu fördern. Somit lässt sich die Steigerung der Studienanfängerquoten erzielen, ohne die Qualitätsmaßstäbe herabsetzen zu müssen.

1.4 Studienabbruch und Studienfachwechsel

Um die Anzahl der Fachkräfte nachhaltig zu steigern, rückt neben der erwähnten Studienberechtigten- beziehungsweise Studienanfängerquote (siehe Kapitel 1.3) insbesondere die Erfolgsquote⁴ der Studienläufe in den Mittelpunkt der Betrachtung. Diese gilt heute als wesentlicher Indikator für das Leistungspotenzial des tertiären Bildungssystems, spiegelt sie doch die Effizienz im Umgang mit Humankapital wider. Demzufolge zeigt eine Erhöhung der Studienanfängerzahlen nur seinen entsprechenden Erfolg für die Volkswirtschaft, wenn die Erfolgsquoten nach der Hochschulexpansion mindestens gleich gehalten, wenn nicht gar gesteigert werden. Um die Erfolgsquoten wiederum zu steigern, muss es gelingen, die richtigen Personen an die Hochschulen zu holen, passende Studiengänge für diese auszuwählen und die Personen dann angemessen zu betreuen, zu fordern und zu fördern.

Eng verknüpft mit dem Begriff der Erfolgsquote sind die Studienabbruch- und die Schwundquote, die nachfolgend genauer erläutert werden:

1.4.1 Studienabbruch – Definitionen und Ausmaße

Studienabbrecher werden in dieser Arbeit verstanden als „ehemalige Studierende, die zwar [...] ein Erststudium an einer deutschen Hochschule aufgenommen haben, dann aber das Hochschulsystem ohne (erstes) Abschlussexamen verlassen und ihr Studium auch nicht zu einem späteren

Alternativen zum Studium zur Verfügung, die in anderen Ländern nicht – oder zumindest nicht in diesem Umfang – vorhanden sind.

⁴ „Erfolgsquoten im Tertiärbereich werden definiert als Prozentsatz derjenigen Studienanfänger im jeweiligen Bildungsbereich, die diesen erfolgreich mit einem ersten Abschluss beenden“ (Organisation für Wirtschaftliche Zusammenarbeit und Entwicklung, 2004, S. 76).

Zeitpunkt wieder aufnehmen“ (Heublein, Schmelzer & Sommer, 2005, S. 9). Die korrespondierende Studienabbrecherquote wird definiert als „Anteil der Studienanfänger eines Studienjahres an den Universitäten und Fachhochschulen, die das Studium beenden, ohne es mit einem Examen in einem Erststudium abzuschließen“ (Heublein, Schmelzer & Sommer, 2005, S. 9). Dabei werden Personen, die das Studienfach und/oder den Studienort wechseln und dann erfolgreich abschließen, nicht als Abbrecher berücksichtigt. Es ist kaum möglich, einen derart definierten Studienabbrecher zweifelsfrei zu identifizieren, da ein temporärer Abbrecher nach Jahren der studentischen Abstinenz an eine Universität zurückkehren könnte (Schröder-Gronostay, 1999). Zur Vereinfachung wird daher meist eine Höchstgrenze an verstrichenen Jahren nach einer Exmatrikulation definiert.

Die im Folgenden berichteten Zahlen zum Studienabbruch stammen primär aus einem verfeinerten Kohortenvergleichsverfahren des Hochschul-Information-Systems Hannover (kurz HIS). Eine umfassende Studienverlaufsanalyse mit Individualangaben ist – wie bei zahlreichen anderen Industriestaaten auch (Hörner, 1999) – zurzeit in Deutschland aus rechtlichen Gründen nicht möglich und muss durch ein Zusammenspiel aus Querschnittsdaten und Absolventenbefragungen ersetzt werden (eine umfassende Beschreibung der Methodik ist in Heublein, Schmelzer & Sommer, 2005 nachzulesen). Durch diese Beschränkung vergeblich Deutschland die Möglichkeit eines notwendigen Bildungscontrollings.

Ausgangsdaten für die folgenden Schätzungen sind die Absolventenzahlen eines Erststudiums, die vom Statistischen Bundesamt (2005a) mit Angabe der Hochschulsemeister gemeldet werden. Aus diesen Daten lassen sich die entsprechenden Studienjahre seit der Erstimmatrikulation berechnen. Der einfachste, aber wenig präzise Weg besteht an dieser Stelle darin, die mittlere Studiendauer in einer bestimmten Fächergruppe und somit ebenfalls den entsprechenden mittleren Studienanfängerjahrgang zu ermitteln. Die Anzahl der Studienabbrecher lässt sich abschätzen, indem die Differenz zwischen den Absolventenzahlen und den Anfängerzahlen im künstlich erzeugten Anfängerjahrgang gebildet wird. Da die Absolventen allerdings nur rechnerisch im Mittel ihr Studium in einem bestimmten Jahrgang begonnen haben – in Wirklichkeit aber zu Teilen auch davor oder danach – werden diese Absolventengruppen im verfeinerten Verfahren den darum gruppierten älteren und jüngeren Anfängerjahrgängen anteilig zugerechnet.

Durch die unterschiedlichen realen Studiendauern eines Absolventenjahrganges können die betrachteten Anfängerjahrgänge stark streuen (zum Beispiel 10 Jahre für das Absolventenjahr 2002, Heublein, Schmelzer & Sommer, 2005). Entsprechend müssen zwischenzeitliche Ereignisse insbesondere auf die mittlere Studiendauer einer Kohorte beachtet werden: So würde beispielsweise ein kontinuierlicher Anstieg der Studiendauer die aus der synthetisierten Kohorte berechnete Abbruchquote ebenfalls ansteigen lassen. Erweitert wird diese rein auf die Daten des Statistischen Bundesamts gestützte Analyse um Befragungen von Studienanfängern (insbe-

sondere um Doppeleinschreibungen quantifizieren zu können) und Absolventen (um Fach- und Hochschulwechsel aufdecken zu können).

Dem oben beschriebenen verfeinerten Kohortenvergleich zufolge verlassen im Durchschnitt 26 % aller Studienanfänger (berücksichtigt sind dabei nur Bildungsinländer) den Tertiärbereich A ohne einen Studienabschluss. Dieser Durchschnittswert variiert natürlich bei einzelnen Fächern und wird insbesondere bei den Sprach- und Kulturwissenschaften (45 %), der Informatik (38 %), den Sozialwissenschaften (36 %) und dem Maschinenbau (34 %) übertroffen (Heublein, Schmelzer & Sommer, 2005). Vergleichsweise geringe Studienabbruchquoten lassen sich in den Fächern Humanmedizin (10 %), Lehramt (12 %), Pharmazie (12 %) und Biologie (15 %) finden. Frauen weisen im Mittel um fünf Prozentpunkte geringere Abbruchquoten als Männer auf.

Um die individuellen Auswirkungen, aber auch den volkswirtschaftlichen Schaden abschätzen zu können, stellt die Studiendauer bis zum Studienabbruch eine wesentliche Größe dar. Im Durchschnitt findet ein Studienabbruch erst nach 7,6 Semestern statt. Der Median liegt bei 5 Semestern (Heublein, Spangenberg & Sommer, 2003). Finden die Studienabbrüche in Mathematik/Naturwissenschaften und Agrar-/Forst- beziehungsweise Ernährungswissenschaften im Durchschnitt nach sechs Hochschulsemestern statt, so zeigen sich bei Kunst und Rechtswissenschaften im Mittel um drei Semester längere Studienzeiten, bevor das Studium abgebrochen wird (Heublein, Spangenberg & Sommer, 2003). Aus individueller wie auch universitärer beziehungsweise staatlicher Sicht sind diese Semester fehlinvestierte Bildungs- und Lebenszeiten mit einem allenfalls marginalen Aspekt der Persönlichkeitsentwicklung. Bei vielen späteren Abbrechern deuten sich schon frühzeitig Probleme im Studienalltag an, die allerdings zu lange verkannt und somit zu spät als ausschlaggebend erkannt werden (Schindler, 1997; G. Sedlacek, 2004).

Bei der weiteren Betrachtung dieser Zahlen muss aber bedacht werden: Studienabbruch ist auch ein Zeichen der Dynamik und Flexibilität des Systems sowie der betroffenen Personen (Lewin, 1999), der erst ab einer bestimmten (nicht näher zu spezifizierenden) Größe in einem Bildungssystem zu einem Indikator für eine generelle Fehlentwicklung wird. Es kann also nicht primäres Ziel sein, die Zahl der Studienabbrecher auf Null senken zu wollen. Dies ist weder aus praktischen, finanziellen, noch aus sozialen und ethischen Gründen umsetzbar. Zweifel sowie Korrekturen an bestimmten persönlichen Entscheidungen nach einer Studienfachwahl in Folge einer kritischen Selbstreflexion und ersten praktischen Erfahrungen mit dem neuen Lernstoff wird und darf es immer geben, auch wenn natürlich den Ursachen durch Studienberatung und -betreuung genauer nachgegangen und – soweit möglich – begegnet werden sollte. Gerade nach der oben beschriebenen Bildungsexpansion stellt der Studienabbruch eben auch eine funktionierende Auslese dar (Gold, 1988; Reissert, 1983), die dann allerdings möglichst frühzeitig – wieder unterstützt durch Betreuungsleistungen – erfolgen sollte.

1.4.2 Schwundquote – Definitionen und Ausmaße

Bei der Interpretation der bis hierher beschriebenen Studienabbrecherquote muss berücksichtigt werden, dass eine Quote von zum Beispiel 20 % keinesfalls bedeutet, dass 80 von 100 Studenten das Studienfach ihrer ersten Wahl erfolgreich abgeschlossen haben. Vielmehr wird lediglich ausgesagt, dass die komplementären 80 % ihr Studium mit einem akademischen Hochschulabschluss (in welchem Fach und nach welcher Studienzeit auch immer) beendet haben. Fachwechsel werden also nicht berücksichtigt, nur der akademische Abschluss steht im Mittelpunkt und wird ausgewertet.

Heublein, Schmelzer und Sommer (2005) verwenden in ihrer Studie daher zusätzlich den Begriff der Studienfachwechsler und bezeichnen damit Personen, die in einem anderen Studiengang als dem zuerst gewählten einen akademischen Abschluss erreichen. Der so genannte Schwund umfasst als Summe von Studienabbrechern und Studienfachwechslern „all jene Studierende eines bestimmten Jahrgangs, die keinen Abschluss in dem Bereich erworben haben, in dem sie sich ursprünglich immatrikulierten“ (Heublein, Schmelzer & Sommer, 2005, S. 9). Bei diesen Berechnungen kann je nach Interesse der Fokus verschoben werden: So werden bei der Kalkulation der Schwundquote eines Studienfaches die Hochschulwechsel nicht berücksichtigt, die Fachwechsel aber schon. Bei der Schwundquote eines Hochschulortes wird hingegen der Fokus auf die Hochschulwechsler verschoben, Fachwechsler werden innerhalb der Hochschule aber nicht beachtet. Im Folgenden werden Schwundquoten – so weit nicht anders bezeichnet – ausschließlich auf Fachebene analysiert. Diese beträgt zum Beispiel in den Sprach- und Kulturwissenschaften 69 %⁵, in den Sozialwissenschaften 68 %, in der Mathematik 65 % und in der Informatik 57 %. Geringe Quoten verzeichnen hingegen insbesondere die Humanmedizin (13 %), Zahnmedizin (19 %), Pharmazie (23 %) und Lehramt (24 %). Das arithmetische Mittel liegt an deutschen Universitäten bei 39 % (Heublein, Schmelzer & Sommer, 2005).

Die bisher berichteten Zahlen bezogen sich lediglich auf Bildungsinländer, also auf Personen die in Deutschland die Hochschulreife erlangt haben. Vergleichbare Studien zu den (Miss-)Erfolgsquoten internationaler Studierender in Deutschland gehen davon aus, dass die Schwundquote (hier allerdings als Konglomerat aus Studienabbruch und Hochschulwechsel zu verstehen) bis zu drei Viertel (Heublein, Sommer & Weitz, 2004) beträgt.

1.4.3 Internationaler Vergleich der Studienabbruchquoten

Auch wenn die bis hierhin berichteten Zahlen zum Studienabbruch und Studienfachwechsel erschreckend hoch erscheinen, liegen sie im internationalen Vergleich im OECD-Mittel. Höhere

⁵ Dies bedeutet, dass von 100 Studienbeginnern in diesem Fachbereich nur 31 Personen einen entsprechenden Studienabschluss erreichen.

Erfolgsquoten als Deutschland (in dieser Studie mit 70 % kalkuliert) verzeichnen insbesondere Japan (94 %), Türkei (88 %), Irland (85 %) und Großbritannien (83 %) (Organisation für Wirtschaftliche Zusammenarbeit und Entwicklung, 2004). Hingegen wurden für Italien (42 % Erfolgsquote), Schweden (48 %) und Österreich (59 %) höhere Abbrecherquoten berechnet. Aufgrund unterschiedlicher Bildungssysteme respektive deren Statistiken sind vergleichenden Interpretationen aber enge Grenzen gesetzt (Hörner, 1999).

ACT berechnet für die USA jedes Jahr die Studienabbruchsquote als so genannte „First to Second Year Retention Rate“ (ACT Inc., 2005), eine so genannte *institutional retention*. Diese Quote stellt rein auf Hochschulebene die Wiederanmeldung (*re-enroll*) auf Jahresebene im Vergleich zu den Anmeldungen im Vorjahr (Seidman, 2005) dar. Dabei werden Hochschulwechsel und Studienabbruch nicht weiter differenziert, beides wird als Misserfolg verstanden. Der Fachwechsel innerhalb einer Institution fällt hingegen nicht auf. Im Falle der Bachelorstudiengänge zeigt sich eine Abbruchquote nach dem ersten Jahr von 33,6 % (an öffentlichen Universitäten) beziehungsweise 29,1 % (an privaten Universitäten). Bei Masterstudiengängen liegen diese Zahlen bei 29,9 % beziehungsweise bei 24,0 %. Ein interessantes Ergebnis zeigt sich in den Vergleichen von Hochschulen unterschiedlicher Selektionsquoten: Öffentliche Colleges, die Bachelor-, Master- und/oder Promotionsstudiengänge anbieten, haben eine durchschnittliche Abbrecherquote zwischen 7,7 % und 37,9 %, moderiert durch die Selektivität des Auswahlprozesses. So gilt die mittlere Abbruchquote von 7,7 % für Schulen, die Personen aufnehmen, die ihren Testergebnissen beziehungsweise den Zeugnisnoten zufolge zu den besten 10 % der Kohorte gehören. Hingegen brechen 37,9 % der Studienanfänger an Hochschulen ab, die als Aufnahmevoraussetzung lediglich einen Highschoolabschluss verlangen. Dabei liegt der mittlere Wert der Erfolgsquoten bei 72,6 % mit einer Standardabweichung von 13,4. Die Zahlen an privaten Colleges schwanken der Studie zufolge zwischen einer Abbrecherquote von 6,5 % und 32,5 %. Die mittlere Erfolgsquote beträgt hier 75,3 % (Standardabweichung 14,0). Erkennbar ist daran, dass der Effekt der Selektivität größer ist als der des allgemeinen Status (öffentlich versus privat) der Hochschule.

Für einige Hochschulen wurden zudem noch die Abbrecherquoten berechnet, die nicht nur den Übergang zum zweiten Jahr berücksichtigen, sondern insgesamt fünf Studienjahre⁶. Dabei liegt die Erfolgsquote zwischen geringen 32,4 % (öffentliche Colleges ohne Zulassungsbeschränkung) und 83,9 % (private Colleges mit einer Zulassung der 10 %-Test- beziehungsweise Notenbesten). Bei diesen Zahlen fällt wiederum die große Varianz auf, so dass Vergleiche mit Deutschland (mittlere Erfolgsquote liegt dabei bei 39 %, siehe Heublein, Schmelzer & Sommer, 2005) nur schwerlich möglich sind.

⁶ Dabei werden Personen auch dann zu den Studienabrechern gezählt, wenn diese erst nach fünf Jahren ihren Bachelorabschluss erhalten. Im direkten Vergleich mit den deutschen Daten der HIS werden also die Abbrecherquoten überschätzt.

Diese Zahlen deuten auf zwei wesentliche Ergebnisse hin: Gerade in öffentlichen US-amerikanischen Colleges zeigen sich – bei aller Vorsicht der Interpretationen – ähnlich geringe beziehungsweise sogar geringere Erfolgsquoten als in Deutschland. Zusätzlich fällt auf, dass Hochschulen mit einer restriktiven Auswahl meist auf der Basis von Studierfähigkeitstests und Zeugnissen auch geringere Abbruchquoten haben. Diese Hochschulen vermögen also effizienter als andere mit dem Gut *Human Resources* umzugehen. Sie wählen vor dem Studium passende Bewerber aus, fordern und fördern diese überaus motivierten Personen und erreichen somit eine wesentliche Grundvoraussetzung für die Spitzenpositionen der amerikanischen Forschungsuniversitäten (Frackmann, 2005).

1.4.4 Auswirkungen von Studienabbruch

Nach der genaueren Betrachtung der zahlenmäßigen Ausmaße des Studienabbruchs und des Studienfachwechsels werden im Folgenden die persönlich-individuellen, volkswirtschaftlichen sowie institutionellen Auswirkungen betrachtet.

Auswirkungen auf das Individuum

Wie Personen subjektiv den eigenen Studienabbruch erleben und einschätzen, hängt von zahlreichen Faktoren ab: so zum Beispiel von den wahrgenommenen Abbruchgründen (insbesondere von der empfundenen Freiwilligkeit des Abbruchs), der Dauer des abgebrochenen Studiums, vom eigenen Alter und von der weiteren Zukunftsplanung (Griesbach, Lewin, Heublein & Sommer, 1998; Schindler, 1999). Je nach Ausprägung dieser Faktoren kann der endgültige Entschluss gegen eine Fortführung des Studiums sowohl als Ausdruck des Humboldtschen Bildungsideals (Spies, 1999), als persönliches Versagen (Gold & Kloft, 1991) oder Versagen des Bildungssystems (Schindler, 1999) ausgelegt werden.

Für die meisten Studienabbrecher waren zweifelsohne ein guter Studienabschluss und die damit verbundenen verbesserten Karrierechancen die zentralen Ziele ihrer Hochschulausbildung (Viebahn, 1990). So ist für viele Betroffene – trotz der positiven Auswirkungen, die auch jede Studienzeit für die persönliche Entwicklung haben kann – die Aufgabe des Studiums ein überaus kritisches Ereignis im Leben (Heublein, Spangenberg & Sommer, 2003). Der endgültige Entschluss zu einem Abbruch ist für viele daher primär gepaart mit dem Eingeständnis der eigenen Fehlentscheidung und des persönlichen Scheiterns. Trotz der hohen Abbrecher- oder Fachwechslerzahlen stellen Studienabbrüche für die meisten eben immer noch eine Abweichung „von der erwünschten Normalität eines erfolgreichen Studiums“ (Schindler, 1997, S. 4) dar. Selbstzweifel und Unsicherheit für die nächsten Karriereschritte, aber auch finanzielle

Probleme sind oftmals Begleiterscheinungen der Neuorientierung. Der vom Individuum wahrgenommene Bruch in der Biografie und in den Karriereplänen hat häufig Auswirkungen bis tief in das eigene Selbstverständnis – mit direkten Folgen auf die weiteren Ausbildungs- oder Beschäftigungsoptionen.

Die objektiven Auswirkungen lassen sich nicht für die gesamte Gruppe der Studienabbrecher generalisieren, zu unterschiedlich sind die betroffenen Individuen, die Beweggründe und somit auch der spätere Karriereweg (Griesbach, Lewin, Heublein & Sommer, 1998). Im Mittel zeigen Studienabbrecher sechs Monate nach der Exmatrikulation aber sowohl eine um ein Drittel höhere Arbeitslosenquote (allerdings liegt diese verglichen mit der Gesamtgesellschaft auf einem insgesamt niedrigen Niveau, siehe Reinberg & Schreyer, 2003) als auch geringere durchschnittliche Bruttoeinkünfte (Heublein, Spangenberg & Sommer, 2003).

Auswirkungen auf die Institutionen

Aus der institutionellen Perspektive betrachtet, blockieren spätere Studienabbrecher, aber unter Umständen auch die meisten Fachwechsler, einen Studienplatz und behindern somit eine straffe Studienorganisation. Ein Studienabbruch betrifft nicht nur das Individuum, sondern eben auch die Institution, also die Hochschule beziehungsweise das Institut, die auf Basis der Erfolgsquote beziehungsweise der Anzahl erfolgreicher Absolventen bewertet werden und teilweise auch dementsprechende finanzielle Zuwendungen erhalten. In den letzten Jahren wurden nämlich die Verteilungsschlüssel zur internen Vergabe des Haushalts auch an deutschen Hochschulen unter anderem um die Berücksichtigung von Absolventen- und Promotionszahlen ergänzt. Somit hat die Verbesserung der Erfolgsquote unter Umständen eine direkte Verbesserung der Haushaltslage auf Institutsebene zur Folge.

Die Erfolgsquote kann aber auch noch eine indirekte Auswirkung auf die Institution haben: In den USA ist jede Hochschule im Rahmen des *student right-to-know and Campus security act* (Public Law 101-542) aus dem Jahr 1990 verpflichtet, die institutionelle Erfolgsquote in Broschüren und auf Internetseiten anzugeben (Seidman, 2005). Diesem Gedanken liegt zugrunde, dass eine Abbruchrate nicht nur von den einzelnen Studierenden abhängt, sondern zu einem hohen Anteil auch von dem durch die Universität geschaffenen Lehr- und Lernumfeld (Qualität der Lehre, Betreuungsleistungen etc.). Mit dieser Veröffentlichung auf Hochschulebene und der daraus resultierenden Vergleichsmöglichkeit ist die Außenwirkung offensichtlich: Gerade angesichts der hohen Kosten, die für ein Studium aufgebracht werden müssen, sollen Studieninteressierte die Möglichkeit haben, mit unterschiedlichen Erfolgswahrscheinlichkeiten zu rechnen und Studienortwahlen anhand dieses Kriteriums zu überdenken. Somit kann die Erfolgsquote als veröffentlichter Wert einen direkten Einfluss auf die kommenden Studienanfängerzahlen und somit auf die Höhe des Etats durch Studiengebühren und staatliche Mittel

ausüben. Inwieweit sich eine Veränderung der Erfolgsquote auch finanziell auf US-amerikanische Hochschulen auswirkt, ist Noel-Levitz (2005) zu entnehmen.

Auswirkungen auf die Volkswirtschaft

Neben den beschriebenen Auswirkungen für die Individuen und für die Institutionen lassen sich auch auf makroökonomischer Ebene Folgen des Studienabbruchs quantifizieren. Zukünftige Studienabbrecher beanspruchen – volkswirtschaftlich gesehen – Gelder, die besser (also mit höherer Rentabilität) hätten investiert werden können. Gerade hinsichtlich der knappen Ressourcen mancher Fachbereiche, der hohen Bewerberzahlen und der damit beklagenswerten Betreuungsrelation erscheint diese Vergeudung nicht tolerabel. Solche Überlegungen und die folgenden Berechnungen gehen vereinfachend davon aus, dass alle von der Volkswirtschaft getätigten Investitionen in einen Studenten durch einen Studienabbruch verloren gehen. Im Bundes- und Fächerdurchschnitt kostet laut Berechnungen der OECD ein Studienplatz pro Jahr etwa 5 500 € (Organisation für Wirtschaftliche Zusammenarbeit und Entwicklung, 2005). Etwa 350 000 junge Leute beginnen pro Jahr ein Hochschulstudium (Statistisches Bundesamt, 2005a), wovon – wie bereits berichtet – 26 % den Tertiärbereich ohne Abschluss (Heublein, Schmelzer & Sommer, 2005) nach durchschnittlich 7,6 Semestern verlassen (Heublein, Spangenberg & Sommer, 2003). Alleine die volkswirtschaftlichen Fehlinvestitionen aufgrund von Studienabbrüchen belaufen sich daher deutschlandweit pro Jahr auf nahezu zwei Milliarden Euro. Dabei sind Fachwechsel noch nicht einmal in die Berechnung einbezogen.

Ein Studienabbruch hat diesen Ausführungen zufolge mehrere volkswirtschaftliche Auswirkungen: Direkt zeigen sich Auswirkungen auf die Akademikerquote, also auf die humanen Ressourcen eines Landes. Somit stellt ein Studienabbruch – zumindest bei Gründen, die eher die Hochschule oder der Staat als das Individuum zu verschulden hat (zum Beispiel durch Probleme der Finanzierung) – eine Verschwendung von gesellschaftlichem und individuellem Humankapital dar. Der Studienabbruch hat aber noch eine weitere volkswirtschaftliche Auswirkung: Durch das medienwirksame Thema des Studienabbruchs werden auch unentschlossene Studienberechtigte beeinflusst und entscheiden sich vielleicht – trotz vorhandener Eignung für ein Studium – gegen die Aufnahme. Gerade strebsame Schüler werden von einem Studium abgehalten und schätzen entsprechend andere, erfolgversprechendere Wege (zum Beispiel eine duale Ausbildung) attraktiver ein.

1.4.5 Gründe für Studienabbruch

Die Gründe für einen Studienabbruch sind so vielfältig wie auch die entsprechenden Begründungen und auch die anvisierten Ziele (Viebahn, 1990) für die Aufnahme eines Studiums. Ein

Studienabbruch ist meist ein komplexer, multikausaler Prozess, der weder ein alleiniges Versagen des Bildungs- oder Hochschulsystems (als Umweltfaktor) noch ein alleiniges Versagen des Individuums (Personenfaktor) ist. Monokausale Begründungen greifen somit klar zu kurz (Schröder-Gronostay & Daniel, 1999), auch wenn die einzelnen Betroffenen häufig einen isolierten Faktor als ausschlaggebend nennen (zum Beispiel eine gerade misslungene Prüfung). In der Mehrzahl der Fälle handelt es sich bei der Entscheidung für einen Studienabbruch um eine Akkumulation von einzelnen Studienabbruchfaktoren. Um im Folgenden die Gründe für einen Studienabbruch beziehungsweise -wechsel genauer beleuchten zu können, werden Personen-, aber auch bestimmte Umweltcharakteristika und deren Interaktionen unterschieden. Dabei lassen sich *Bedingungsfaktoren* und *Studienabbruchmotive* (Heublein, Spangenberg & Sommer, 2003) differenzieren: Die einzelnen *Bedingungsfaktoren* moderieren die Wahrscheinlichkeit eines Studienabbruchs. Einteilen lassen sich diese in individuelle (Intelligenz, psychische Stabilität, Leistungsmotivation, Volition etc.), institutionell-gesellschaftliche Faktoren (Betreuung vor und während des Studiums, Studienbedingungen etc.) und deren Interaktionen. Die einzelnen Bedingungsfaktoren lösen nach Heublein, Spangenberg und Sommer (2003) nicht unmittelbar den Abbruch aus, sondern „beeinflussen die innere Motivationslage“ (S. 7) und können damit als Ursache für die eigentlichen *Studienabbruchmotive* gelten.

Studienabbruchmotive

Die unterschiedlichen Abbruchmotive mit den jeweiligen Häufigkeiten werden im Folgenden vorgestellt: Im Rahmen einer aktuellen repräsentativen Studie des Hochschul-Informationssysteme HIS (Heublein, Spangenberg & Sommer, 2003) gaben 17 % der befragten Studienabbrecher an, aus *finanziellen Gründen* das Studium vorzeitig beendet zu haben. Bei einem ähnlich hohen Anteil war der Wunsch nach einer *beruflichen Neuorientierung* ausschlaggebend. Von ihnen werden demnach nicht mehr ein theorielastiges Studium und ein entsprechender Beruf präferiert, sondern meist eine Berufsausbildung, in Verbindung mit praktischeren Tätigkeiten. 16 % der Befragten führten eine *fehlende Studienmotivation* als ausschlaggebend an. Dabei stehen Desinteresse am möglichen Beruf, nachgelassenes Interesse an den Inhalten und falsche Erwartungen an das Studienfach im Mittelpunkt. Eines dieser drei Abbruchmotive wählten 50 % der befragten Abbrecher. Seltener wurden Leistungsprobleme (11 %; zum Beispiel „Anforderungen zu hoch“, „Leistungsdruck“), familiäre Probleme (10 %; „Studium und Kinderbetreuung nicht vereinbar“, „Schwangerschaft“), Prüfungsversagen (8 %; „Abschlussprüfung nicht bestanden“) und eine eigene Krankheit (5 %) als Grund angegeben. Dabei zeigen sich große Unterschiede zwischen den Fächern: So treten Leistungsprobleme bei den Studienfächern Informatik/Mathematik (19 %), Ingenieurwissenschaften (19 %) und Medizin (14 %) überdurchschnittlich häufig auf, hingegen kaum beim Fach Kunst (5 %). Prüfungsversagen erklärt den Studienabbruch bei den Rechtswissenschaften gar zu 26 %, aber bei den Sprach-/Kulturwissenschaften beziehungsweise Sport lediglich zu 2 %.

Augenscheinlich ist, dass sich ein großer Anteil der Personen aufgrund einer empfundenen Distanz zum Studium, zum wissenschaftlichen Arbeiten, zum Lernen etc. beziehungsweise aufgrund der resultierenden Studienergebnisse entschließt, das derzeitige Studium abzubrechen. Somit lässt sich vermuten, dass die anfängliche Entscheidungsfindung, die zur Studienfachwahl führte, häufig nicht ausreichend reflektiert sowie Konsequenzen und Alternativen nicht ausreichend bedacht worden sind (Heine & Willich, 2006). Zu häufig bestehen unrealistische Vorstellungen, Einschätzungen und Erwartungen über die zu bewältigenden Anforderungen und die dazu erforderlichen Fähigkeiten und Eignung, kurz: die für eine rationale und nachhaltige Entscheidungsfindung erforderliche Kompetenz ist bei einem Teil der Jugendlichen und jungen Erwachsenen nur unzureichend vorhanden.

Die einzelnen Häufigkeiten, mit denen die benannten Studienabbruchmotive zu einem Studienabbruch führen, können der Grafik 1.1 entnommen werden. Bei den Prozentangaben werden zwischen frühen (erste Prozentzahl; Abbruch nach einer Studiendauer von maximal zwei Semestern) und späten Studienabbrechern (zweite Prozentzahl, Studiendauer über 2 Semester) unterschieden.

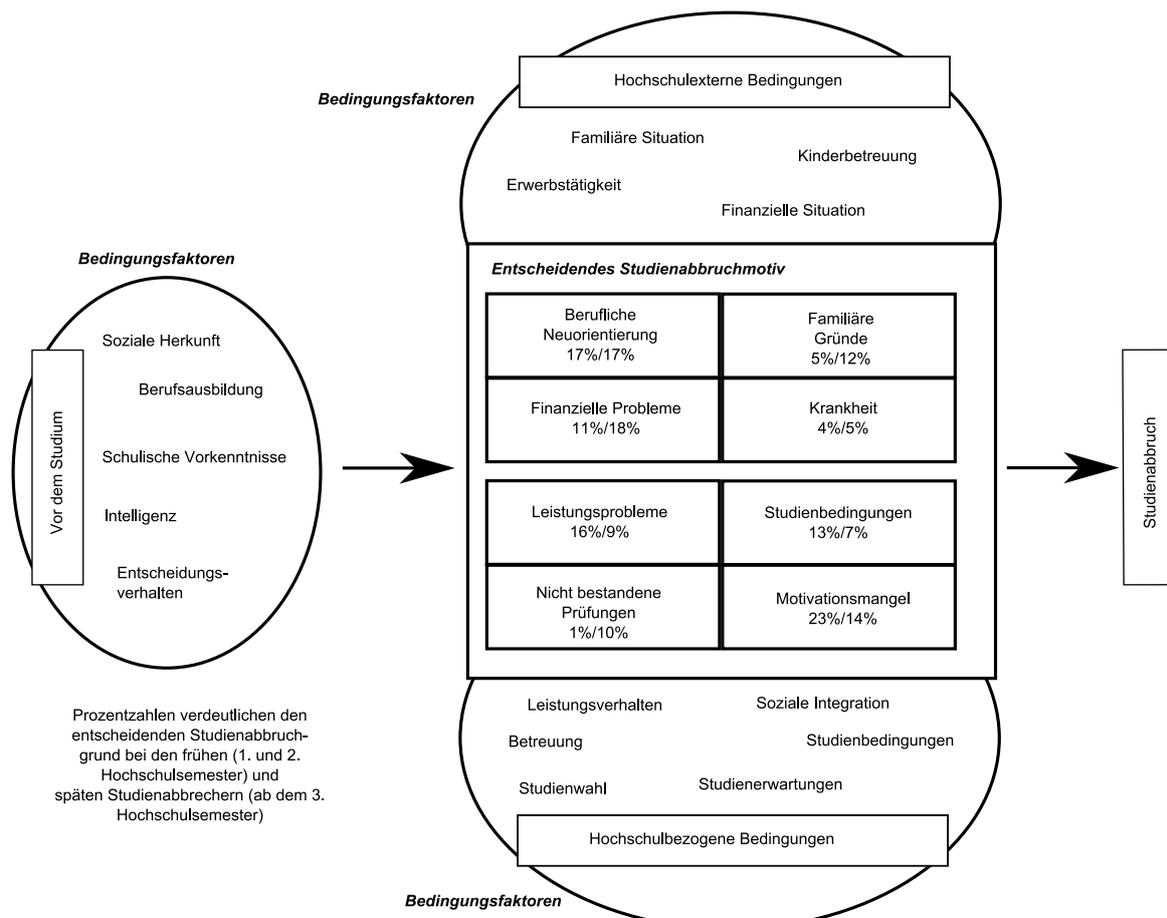


Abbildung 1.1: Studienabbruchmotive und deren Bedingungsfaktoren, stark modifiziert und erweitert nach Heublein, Spangenberg & Sommer, 2003

Bedingungsfaktoren

Wurden im vorigen Abschnitt primär die Studienabbruchmotive genannt, folgen nun die Bedingungsfaktoren, die zwar subjektiv nicht als ausschlaggebend wahrgenommen werden, aber einen Studienabbruch wahrscheinlicher machen.

Eine wesentliche Einflussgröße im deutschen Bildungssystem stellt die *soziale Herkunft* dar, die nicht nur einen allgemeinen Einfluss auf die Lese- und Mathematikkompetenz (Prenzel et al., 2004) hat, sondern eben indirekt auch auf die Wahrscheinlichkeit, die Allgemeine Hochschulreife zu erreichen und ein Studium aufzunehmen (Bargel, 2005; Georg, 2005; Isserstedt, Middendorff, Weber, Schnitzer & Wolter, 2004; Organisation für Wirtschaftliche Zusammenarbeit und Entwicklung, 2004; Simeaner, Röhl & Bargel, 2004). Auch im internationalen Vergleich mit anderen industrialisierten Ländern zeigen sich in Deutschland Kinder aus bildungsferneren Schichten unter den Studierberechtigten und Studienanfängern unterrepräsentiert (Isserstedt, Middendorff, Weber, Schnitzer & Wolter, 2004; Lewin, Heublein, Schreiber, Spangenberg & Sommer, 2002). Über einen Zeitraum der letzten zwei Jahrzehnte hinweg betrachtet, hat sich der Anteil der Studierenden aus der Herkunftsgruppe mit hohem sozialen Stand⁷ sogar nahezu verdoppelt. Im Gegenzug reduzierte sich der Anteil Studierender aus bildungsfernen Milieus auf fast die Hälfte, ohne dass eine vergleichbare Tendenz auch allgemein für Deutschland nachweisbar ist (Isserstedt, Middendorff, Weber, Schnitzer & Wolter, 2004). Diese soziale Selektion setzt sich an den Hochschulen fort, indem Kinder aus bildungsfernen beziehungsweise statusniedrigeren Elternhäusern länger studieren und häufiger ihr Studium aufgeben (Isserstedt, Middendorff, Weber, Schnitzer & Wolter, 2004; Heublein, Spangenberg & Sommer, 2003) müssen. Der Einfluss des sozialen Standes der Eltern auf den Studienverlauf des Kindes ist aber nicht nur in Deutschland augenscheinlich. So lässt sich eine entsprechende Abhängigkeit der Abbrecherquote vom sozialen Stand der Familie auch in den USA nachweisen (Tinto, 1993). Dabei liegt die Erfolgsquote von Personen mit einem sozioökonomischen Status im oberen Quartil nach sechs Jahren Collegezeit bei 55,4 %, von Personen mit einem Status im unteren Quartil hingegen nur bei 30,1 %.

Eng mit der sozialen Herkunft ist die *Studienfinanzierung* verknüpft: Studenten mit einem niedrigen sozialen Stand der Eltern arbeiten durchschnittlich häufiger, regelmäßiger als Personen mit einer höheren sozialen Herkunft (Isserstedt, Middendorff, Weber, Schnitzer & Wolter, 2004) und mit einer anderen Motivation, nämlich verstärkt zur Selbstfinanzierung und weniger um einen verstärkten Konsum zu ermöglichen. Zusätzlich steigt die Eigenfinanzierung der Studenten mit sozial niedriger Herkunft besonders nach dem Ende der Finanzierung nach dem BAFöG an, die meist an die Regelstudienzeit gekoppelt ist (Heublein, Spangenberg & Sommer, 2003). Hingegen investieren Studierende aus hochschulnahem Elternhaus eher mehr Zeit in der

⁷ Bei der Bestimmung des sozialen Standes wird die berufliche Stellung, der höchste berufsqualifizierende und der höchste allgemeinbildende Abschluss der Eltern betrachtet.

Abschlussphase ihres Studiums als zuvor (Isserstedt, Middendorff, Weber, Schnitzer & Wolter, 2004). Studenten, die auf Eigenfinanzierung angewiesen sind, können dies hingegen nicht. Erwartungsgemäß zeigt sich, dass der individuelle Studienverlauf von der Höhe zur Verfügung stehender finanzieller Mittel und der Art ihres Erwerbs maßgeblich beeinflusst wird: Dabei ist bei längerer Erwerbstätigkeit pro Woche sowohl ein Absinken der Anzahl bestandener Prüfungen als auch ein Anstieg der Studiendauer und auch eine Verschlechterung der Prüfungsnoten beobachtbar (Brandstätter & Farthofer, 2003a). Entsprechend sind auch überproportional häufig finanzielle Probleme für einen Studienabbruch entscheidend (Heublein, Spangenberg & Sommer, 2003; Reissert, 1991). Dabei scheint jedoch insbesondere die Quelle der Finanzierungsmöglichkeit und deren Rahmenbedingungen (Arbeitszeiten, Art der Tätigkeit) wesentlich zu sein. Die besondere Bedeutung der staatlichen Finanzierungsmöglichkeiten (vornehmlich nach dem BAFöG) wird durch diese vorgestellten Ergebnisse offensichtlich, ist doch erwünscht, dass geeignete Personen unabhängig von der sozialen Herkunft und Finanzierungsmöglichkeiten studieren.

Ein dritter wesentlicher Bedingungsfaktor ist in der unterschiedlichen *schulischen Vorbildung* (als Vorwissen für ein Studium zu verstehen), aber auch in stabilen *Fähigkeiten* und *Persönlichkeitseigenschaften* zu finden. Eine erfolgreiche Bewältigung eines Studiums gelingt empirischen Ergebnissen nach (Deidesheimer Kreis, 1997; Heublein, Spangenberg & Sommer, 2003; Konegen-Grenier, 2001) insbesondere dann, wenn bei den Studierenden sowohl die passenden Fach- und Methodenkompetenzen als auch die Fähigkeiten zur Kommunikation und zur Selbstorganisation vorhanden sind (siehe Abschnitt 2.2.1). Ein mangelndes Leistungsvermögen wirkt sich häufig allerdings nicht erst auf den Lernerfolg an der Universität, sondern bereits viel früher beim Erwerb der Basiskenntnisse an den Schulen aus. Bereits 1981 im Rahmen des Studie Bildungsläufe (Giesen et al., 1981) konnte nachgewiesen werden, dass Abbrecher und Wechsler einen schlechteren Notendurchschnitt in der Schule hatten als erfolgreiche Studierende. Dies gilt auch heute: Nahezu 60 % der erfolgreichen Absolventen haben die Schule mit einer Abschlussnote „gut“ oder „sehr gut“ abgeschlossen. Die entsprechende Quote liegt bei den späteren Abbrechern um etwa 20 Prozentpunkte niedriger (Heublein, Spangenberg & Sommer, 2003). Demgemäß nennen rund ein Viertel der Studienabbrecher zu hohe Studienanforderungen als einen wichtigen Grund für einen Abbruch. Erhebliche Diskrepanzen zwischen Abbrechern und Absolventen lassen sich sowohl hinsichtlich der mathematischen und naturwissenschaftlichen Vorkenntnisse (insbesondere in den Fächergruppen Mathematik/Naturwissenschaften, Wirtschafts- beziehungsweise Sozialwissenschaften und Ingenieurwissenschaften) als auch hinsichtlich der Fähigkeit zur selbstständigen Studiengestaltung (insbesondere in Rechtswissenschaften, Kunst, den Lehramtsstudiengängen und in Sprach-/Kulturwissenschaften) finden (Heublein, Spangenberg & Sommer, 2003). Zu Studienbeginn fehlten einer Befragung zufolge 29 % der späteren Studienabbrecher bzw. 18 % der späteren Absolventen in hohem Maße mathematische Vorkenntnisse. 14 % beziehungsweise 8 % hatten nach eigenen Angaben nur unzureichendes

naturwissenschaftliches Wissen. 26 % der ohne Abschluss Exmatrikulierten, aber nur 18 % der Absolventen hatten bei Studienbeginn nicht die Fähigkeit zur selbstständigen Studiengestaltung.

Der vierte wesentliche Bedingungsfaktor beschäftigt sich mit der *Studienentscheidung* selbst. Die Entscheidung für ein Studienfach oder für einen bestimmten Beruf gilt als eine der zentralen Lebensentscheidungen, die den beruflichen und häufig auch privaten Weg für einen Zeitraum der nächsten Jahrzehnte bereitet. Erschreckend häufig zeigt sich – trotz dieser Bedeutung – in Gesprächen mit Studienabbrechern, dass deren Studienentscheidung von einer großen Ahnungslosigkeit bestimmt war (zur genaueren Betrachtung des Entscheidungsprozesses siehe Abschnitt 2.1): Studieninhalte und -anforderungen wie auch berufliche Perspektiven waren vielfach nicht bekannt. Nur etwa ein Drittel aller Studienanfänger in Deutschland ist eigenen Auskünften zufolge gut über das gewählte Studium und die Hochschule informiert (Lewin, Heublein, Schreiber, Spangenberg & Sommer, 2002). Ähnliche Ergebnisse zeigen US-amerikanische Studien (Greening, 1998; Tinto, 1993).

Die schlechte Informationslage zeigt direkte Auswirkungen auf die individuelle Erfolgsquote: Bei jedem zweiten Studienabbrecher in Deutschland haben falsche Erwartungen als „Distanz zu den fachspezifischen Inhalten“ (Apenburg, 1979; Lewin, 1999, S. 29) eine bedeutende Rolle bei der Abbruchentscheidung gespielt. Bei ihnen haben sich die Vorstellungen zu Studienbeginn seltener eingelöst als bei den Absolventen (Heublein, Spangenberg & Sommer, 2003): Immerhin 44 % der Absolventen sehen ihre damaligen Erwartungen alles in allem erfüllt, aber lediglich 18 % der Abbrecher. Im Mittel entscheiden sich spätere Wechsler und Abbrecher zudem seltener für einen bestimmten Beruf (den sie nach dem Studium gerne ausüben möchten) und treffen die Studienfachwahl vergleichsweise spät (Giesen et al., 1981). Analog konnte an US-Colleges aufgezeigt werden, dass 80 % der späteren Studienabbrecher ihre wesentliche Karriereentscheidungen bei der Studienfachentscheidung noch nicht abgeschlossen hatten und sich so entsprechend unsicher zeigten (Waterman & Waterman, 1972). Ein damit verwandtes Problem: Zu häufig werden Studienfachwahlen auf der Basis günstiger Arbeitsmarktaussichten (Heublein, Schmelzer, Sommer & Spangenberg, 2002) getroffen, nicht aber auf Basis der Studieninhalte, der eigenen Interessen und der Fähigkeiten. Eine einseitige Informationslage (spätere Einkommen, besonderer Fachkräftemangel etc.) und eine klare Empfehlung ersetzt so zu häufig die Auseinandersetzung mit einem Studienfach. Angeheizt durch reißerische Artikel wie zum Beispiel „Gute Fächer, schlechte Fächer“ (Koch & Mohr, 2006) des Wochenmagazins „Der Spiegel“ vernachlässigen Schulabsolventen die persönlich-individuelle Dimension des Entscheidungsprozesses. Der Arbeitsmarkt ist ein wichtiges, er darf allerdings kein alleiniges Entscheidungskriterium für die Fachwahl sein. Im Abschnitt 2.2.1 wird das entscheidende Konzept der Personen-Umwelt-Passung genauer vorgestellt, das hier als Ausweg dienen kann.

Institutionell-gesellschaftliche Einflussgrößen

Etwa bei drei Viertel der Studienabbrecher hat mindestens ein unzulänglicher Aspekt der Studienbedingungen (zum Beispiel „Lehrkräfte waren nicht immer gut vorbereitet“, „Lehrveranstaltungen litten an Überfüllung“ etc.) eine wichtige Rolle bei der Abbruchentscheidung gespielt (Heublein, Spangenberg & Sommer, 2003). Etwa ein Zehntel der ohne Abschluss Exmatrikulierten ist an den Studienbedingungen selbst gar gescheitert.

Interaktion von individuellen und institutionellen Einflussgrößen

Die bis hierhin beschriebenen Einflussgrößen beschäftigen sich primär entweder mit institutionellen oder aber mit individuellen Besonderheiten. Im Folgenden steht das Zusammenspiel dieser beiden Faktoren verstärkt im Mittelpunkt. So scheint der Grad der sozialen Integration in der Hochschule ein weiterer wesentlicher Bedingungsfaktor und somit ein entscheidender Prädiktor für ein erfolgreiches Studium zu sein. Bereits Anfang der 1970er Jahre formulierten Spady (1970) und darauf aufbauend Tinto (1993) für das US-amerikanische Hochschulsystem den Zusammenhang zwischen sozialen Beziehungen und dem freiwilligen Studienabbruch. Entstanden sind diese Ansätze aus Durkheims Theorie (1973) zum Suizid. Dabei wird ein Studienabbruch – analog dem persönlichen Bruch mit dem normativen und sozialen System einer Gesellschaft und dem folgenden Selbstmord – mit einer defizitären sozialen Integration in das Hochschulsystem erklärt. Für die genauere Betrachtung der sozialen Integration rücken dabei die persönlichen Interessen und Zielsetzungen einerseits und die institutionellen Erwartungen und Einflüsse andererseits in den Mittelpunkt. Je mehr diese sich entsprechen, desto eher ist eine erfolgreiche Assimilation an das soziale und an das akademische System des College oder eben übertragen der deutschen Hochschule zu erwarten. Die Assimilation zeigt sich entsprechend durch feste Freundschaften beziehungsweise dem Zusammenleben mit anderen Studierenden, durch anregende Mitarbeit in studentischen Lerngruppen und Anerkennung durch Dozenten. All diese Faktoren beeinflussen sowohl den Studienerfolg als auch die Studienzufriedenheit positiv und senken indirekt auch die Wahrscheinlichkeit eines Studienabbruchs (Beil, Reisen, Zea & Caplan, 1999; Elkins, Braxton & James, 2000; Pascarella & Chapman, 1983; Stoecker, Pascarella & Wolfe, 1988).

Tintos Interaktionstheorie gilt besonders im angloamerikanischen Bereich als wichtigstes Modell zum Studienabbruch, lässt sich aber nur eingeschränkt auf das deutsche Hochschulsystem übertragen (Gold, 1988). Entsprechend sind zahlreiche US-amerikanische Publikationen dazu aufzufinden, wenige aber nur zu europäischen Bildungssystemen (z.B. Gesk, 1999). Interessierte Leser seien insbesondere auf Braxton, Sullivan und Johnson (1997) verwiesen, die das Modell anhand von 15 prüfbaren Aussagen zusammenfassen und anschließend eine Literatursichtung

durchführen, um die Aussagen empirisch zu bestätigen. Dabei konnten lediglich sieben der 15 Aussagen eindeutig durch Studien gestützt werden.

1.5 Zusammenfassung und Ausblick

In diesem Kapitel wurde versucht, die Bedeutung eines international wettbewerbsfähigen Bildungssystems zu verdeutlichen: Deutschlands Zukunft als innovatives Industrieland ist eng mit einer langfristigen Sicherung und Vergrößerung des Humankapitals verbunden. Nicht nur Qualität, sondern eben auch die Quantität von Fachkräften ist im weltweiten Wettbewerb entscheidend. Dabei muss der scheinbare Gegensatz von Qualität und Quantität gelöst und beide Kennwerte gleichermaßen gesteigert werden: Die Fokussierung auf die Seite der Quantität, also möglichst viele Personen an die Hochschulen zu locken – ungeachtet ihrer Eignung – kann nicht Ziel einer verantwortungsvollen Bildungspolitik sein. Aber auch eine weitere Elitisierung ohne eine Berücksichtigung der breiten, durchaus geeigneten Masse würde dem Bildungsstandort mehr schaden als nützen, weil eben nicht nur wenige hervorragend, sondern auch viele gut ausgebildet werden müssen.

Ein wesentlicher Ansatzpunkt ist die Verbesserung der universitären Erfolgsquoten. Eine Volkswirtschaft kann und darf es sich nicht leisten, humane Ressourcen durch Studienabbrüche und Fachwechsel zu vergeuden. Frühzeitig müssen Schulabsolventen unterstützt werden bei der Beantwortung zentraler Fragen rund um die Studienfachwahl. Was bedeutet es, das Fach „Informatik“ zu studieren? Was sind da überhaupt die Studieninhalte? Würde mir das Spaß machen? Bin ich dazu geeignet? Kann ich eigentlich die mathematische Leistung erbringen, die dort von mir gefordert wird?

Zur Beantwortung dieser Fragen reicht es nicht, lediglich allgemeine Informationen über den Arbeitsmarkt zur Verfügung zu stellen. Vielmehr muss die individuelle Passung der Person und des Studienfaches betrachtet werden. Das jeweils optimal passende Fach muss ausgewählt werden, um eine hohe Erfolgswahrscheinlichkeit erreichen zu können. Im Mittelpunkt der Ausführungen im nächsten Kapitel werden daher zwei unterschiedliche Wege zur Studienfachwahl genauer beschrieben. Gegenübergestellt wird dabei die irrationale Entscheidungsfindung und eine Wahl auf Basis eines differenzierten Eignungs- und Neigungsbegriffs.

Kapitel 2

Eignungsbeurteilungen in der Fremd- und Selbstselektion von Studieninteressierten

Im vorigen Kapitel wurde die Irrationalität vieler Studienfachwahlen angesprochen: Fachwahlen werden häufig überstürzt auf der Basis einfach verfügbarer Informationen vorgenommen – nicht selten mit erschreckenden Resultaten und geringen Erfolgsquoten. In diesem Kapitel wird der irrationale Entscheidungsprozess genauer beleuchtet, aber auch eine rationale Vorgehensweise auf Basis der eigenen Eignung und Neigung entworfen.

2.1 Die irrationalen Entscheidungen bei einer Studienfachwahl

Jeder Entscheidungsfindung hinsichtlich der Karriere und somit auch jeglicher Studienfachwahl liegen vier generelle Elemente zugrunde, die von den Entscheidern unterschiedlich umfassend berücksichtigt werden: Zentral sind insbesondere die *selbst gesetzten Ziele*, die es durch ein Studium zu erreichen gilt (sichere Arbeitsmarktchancen, kurze Studiendauer bis zum Studienabschluss, geringe Kosten, Befriedigung der studienfachspezifischen Neugierde etc.; siehe Abschnitt 3.1), die zur Verfügung stehenden *Wahlalternativen* (also die einzelnen Studienfächer), die im riesigen Angebot erkannt werden müssen, *deren wahrscheinlichsten Auswirkungen* (Studienerfolgsquote, mögliches Einkommen, Aufstiegschancen, Arbeitszeiten etc.) und die resultierenden *Bewertungen dieser Auswirkungen* auf Basis der selbst definierten Ziele (Pitz & Harren, 1980). Umschrieben wird diese auf die berufliche Karriere ausgerichtete Entscheidungsfindung mit dem Begriff *career decision-making*.

Die bereits erwähnte Irrationalität zum Beispiel in der Studienfachwahl lässt sich anhand dieser vier Bestimmungselemente umschreiben als ein Entscheidungsprozess, der – anders als ein rationales Vorgehen – nicht für Außenstehende nachvollziehbar und transparent ist. So werden

in diesen Fällen keine oder nur unklare Ziele definiert, die Auswirkungen nicht analysiert und die Schlussbewertung nicht auf der Basis der vorgenannten Informationen vollzogen. Dabei ist wichtig zu erwähnen, dass auch eine rational getroffene Entscheidung für Außenstehende nicht zwangsläufig sinnvoll erscheinen muss. Rational ist eine Entscheidungsfindung immer schon dann, wenn Außenstehende bei Berücksichtigung derselben Alternativen, Ziele und Auswirkungen dasselbe Studienfach empfehlen würden.

Im Folgenden werden die einzelnen Phasen und die damit verbundenen Schwierigkeiten in der Studienfachwahl genauer betrachtet:

Nach dem Erhalt der Allgemeinen Hochschulreife stehen den Absolventen im deutschen Bildungssystem sowohl der Sekundar- (somit eine Ausbildung im dualen System) als auch der Tertiärbereich (ein Fach- oder Hochschulstudium) offen. Nicht nur der Entschluss für ein Hochschulstudium und damit gegen eine Ausbildung im dualen System stellt für die meisten Schüler ein diffiziles Problem dar. Auch nach einer erfolgten Entscheidung für den Tertiärbereich stehen derzeit an deutschen Hochschulen etwa 9 000 grundständige Studiengänge als Wahlmöglichkeiten zur Verfügung (Hochschulrektorenkonferenz, 2006) – Tendenz steigend.

Um die dabei in engere Wahl genommenen Alternativen bewerten zu können, fehlt es häufig an einer genauen persönlichen Zieldefinition für das Studium. Was soll durch ein Studium erreicht werden? Soll das Studium interessant sein oder reicht es, ein Studium zu wählen, dass gute Berufsaussichten beschert? Ziele werden oftmals unklar definiert – wenn überhaupt. Für die meisten Studieninteressenten besteht das wichtigste Ziel – zumindest entsprechend einer Selbstauskunft – darin, ein Studienfach zu finden, dass interessante Sachverhalte vermittelt (Heublein & Sommer, 2002; Simeaner, Röhl & Bargel, 2004). Diese Einschätzung gilt rückblickend für spätere Absolventen häufiger (87 %) als für spätere Studienabbrecher (78 %) (Heublein, Spangenberg & Sommer, 2003). Letztere haben also signifikant seltener das Fachinteresse als Ziel in den Entscheidungsprozess aufgenommen. Mehr als 40 % der Abbrecher und 33 % der Absolventen führen extrinsische Motivationsquellen an, die zu der Studienfachentscheidung geführt haben. Diese Quellen haben gemein, dass nicht die Auseinandersetzung mit einem Gegenstand selbst motiviert, sondern Abwägungen des Nutzens (späterer Verdienst, Ansehen, günstige Berufsaussichten) einer Studienfachwahl.

Nach dem Auffinden der Alternativen und der Festlegung der Ziele müssen die Alternativen genauer analysiert werden. Gerade dabei zeigt sich, dass die Vielzahl der Möglichkeiten kaum zu überblicken ist. Dadurch scheint es besonders notwendig, Studieninteressierten eine Hilfe zu bieten, indem die Fächerflut nach eigenen Wünschen – zum Beispiel nach den Interessen (wenn Fachinteresse als Ziel definiert ist) – gefiltert werden kann. Ein solches System würde ermöglichen, dass die Studieninteressierten eine größere, bereits gefilterte Übersicht über die Alternativen erhalten. Aber auch die Analyse von Informationen über die Studienfächer ist nur ein Teil, fehlen doch noch weitere Auswertungen über die unterschiedlichen Wahlmöglichkeiten,

die unter anderem auch die Interaktion von Person und Studienfach umfassen (Germeijs & Boeck, 2003; Pitz & Harren, 1980): Werden die Studieninhalte, wird die Studienumwelt zur Person passen? Wie wird sich die Studienfachwahl auswirken auf die Zufriedenheit, wie auf die Berufschancen?

Bei allen Analysen zur Entscheidungsvorbereitung werden von Menschen nicht nur relevante, sondern auch unpassende, aber schnell verfügbare Informationen gesammelt (Harren, 1979). So werden Entscheidungen häufig auf Basis falscher Informationen, verzerrter Bilder oder Entscheidungskriterien gefällt, die wenig bis überhaupt keine Beziehungen zu den Inhalten des späteren Studienfachs haben. Zum Beispiel beeinflussen Fach- beziehungsweise Berufsstereotype (Hannover & Kessels, 2004; Höttecke, 2001; Zwick & Renn, 2000) genauso die Entscheidung wie das soziale Umfeld des eigenen Elternhauses oder des Freundeskreises (Simeaner, Röhl & Bargel, 2004; Wissenschaftsrat, 2004), positive Arbeitsmarktaussichten (Heine, Bechmann & Durrer, 2002; Heublein & Sommer, 2002), die Nähe zur Familie beziehungsweise die Attraktivität des Hochschulstandortes (Engelhardt, 1993; Stemmler, 2005; Wissenschaftsrat, 2004) oder der Zufall (Maurice, Scheller & Bäumer, 1995).

Die Ausführungen in diesem Abschnitt legen nahe, dass die Begründungen für überstürzte Entscheidungen häufig nicht nur in den Personen selbst zu suchen sind. Vielmehr sind die meisten Studieninteressierten schlichtweg ohne professionelle Hilfe nicht in der Lage, die persönlichen Ziele für ein Studium festzulegen, die Alternativen zu analysieren und sich auch selbst einzuschätzen. Zu schwierig scheint der Prozess der wirklich objektiven Studienfachwahl. Genau in diesem Entscheidungsprozess muss mehr als bisher unterstützt werden, eine möglichst optimale Entscheidung zu treffen. Die reine Gabe von Informationen zum Studium oder gar nur zum Berufsfeld und dessen Arbeitsmarktchancen greift klar zu kurz. Wichtig erscheint es vielmehr, die Passung zwischen den Interessenten und den Fächern zur Grundlage einer umfassenderen Platzierung zu machen. So wird es möglich sein, die durchschnittliche Erfolgsquote zu steigern und somit das vorhandene Humankapital in Deutschland effizienter zu nutzen. Im folgenden Kapitel wird das Thema der Passung zwischen Studienfach und Person und deren Auswirkungen umfassender vorgestellt.

2.2 Die rationale Studienfachwahl auf Basis der Eignung und Neigung

2.2.1 Die Person-Studienfach-Passung

In dem vorangegangenen Kapitel wurde betont, dass Personen unterschiedlichste Alternativen, deren Auswirkungen und die schlussendliche Einzelbewertung der Alternativen bedenken müssen. Jeder Entscheider kann dabei selbständig auswählen, nach welchen mehr oder weniger sinnvollen Kriterien die Studienfachwahl vorgenommen wird. Im Folgenden wird das Eignungskonzept als Passung zwischen Person und Beruf beziehungsweise Studienfach näher erläutert. Nach diesem Konzept ist es förderlich für den Studienverlauf, wenn das Studienfach nach den eigenen Fähigkeiten, Fertigkeiten und Interessen ausgewählt wird und nicht zum Beispiel nach den Verdienstmöglichkeiten oder dem Ansehen des Studienfaches (Amelang, 1997; Deidesheimer Kreis, 1997).

Frank Parsons postulierte bereits im Jahr 1909 in seinem posthum veröffentlichten Buch *Choosing a Vocation* eine Berufswahltheorie, die den Berufswahl- oder auch Studienwahlprozess auf der Basis einer Passung zwischen Person und Umwelt beziehungsweise zwischen Student und Studienfach beschreibt (Parsons, 1909). Diese Publikation gilt als Ausgangspunkt der heutigen Berufsberatung (Dixon, 2000). Parsons skizziert den ersten Schritt einer jeden Karriereentscheidung als eine Phase, in welcher der Unentschiedene mehr über die eigene Eignung und Neigung in Erfahrung bringt. Das Ziel dieser ersten Phase ist es somit, ein klareres Bild über sich, über die eigenen Fähigkeiten, als auch über Wünsche und Bedürfnisse zu erhalten. Neben dieser Verfeinerung beziehungsweise Korrektur des beruflichen Selbstkonzepts¹ streben die Personen in einer zweiten Phase an, mehr über die verschiedenen Ausbildungs- und Berufswahlmöglichkeiten zu erfahren. Im abschließenden Schritt muss die Person Beziehungen zwischen den eigenen Merkmalen und denen der Wahlmöglichkeiten herstellen und die resultierende Passung abschätzen.

Das beschriebene Modell der Passung zwischen Beruf und den Fähigkeiten, Fertigkeiten und Interessen lässt sich nicht nur im Bereich der Berufswahl, sondern auch im Bereich der Studienfachwahl anwenden. Dabei wird – analog zu Apenburg (1980) – ein Beruf einem Studienfach gleichgesetzt. Entsprechend lässt sich diese Form der Passung – übertragen aus der Forschung zur Arbeitswelt (Weinert, 1992) – auch bei der Betrachtung der Eignung für ein bestimmtes Studienfach (Amelang, 1997) *Personen-Umwelt-Passung* nennen (oder auch

¹ Nach Herkner (1991) ist das Selbstkonzept das Wissen, dass eine Person über sich selbst hat. Dies umfasst Informationen „über das Aussehen, Verhaltensweisen und Ergebnisse (Erfolge, Misserfolge) in verschiedenen Situationen, Eigenschaften, Einstellungen und Ziele (Motive). Außerdem enthält es Wissen über die sozialen Beziehungen (Freunde, Familie, Kollegen, usw.), Kategorienzugehörigkeiten (Geschlecht, Beruf, usw.) und über innere Prozesse wie Gefühle und Stimmungen“ (Herkner, 1991, S. 361).

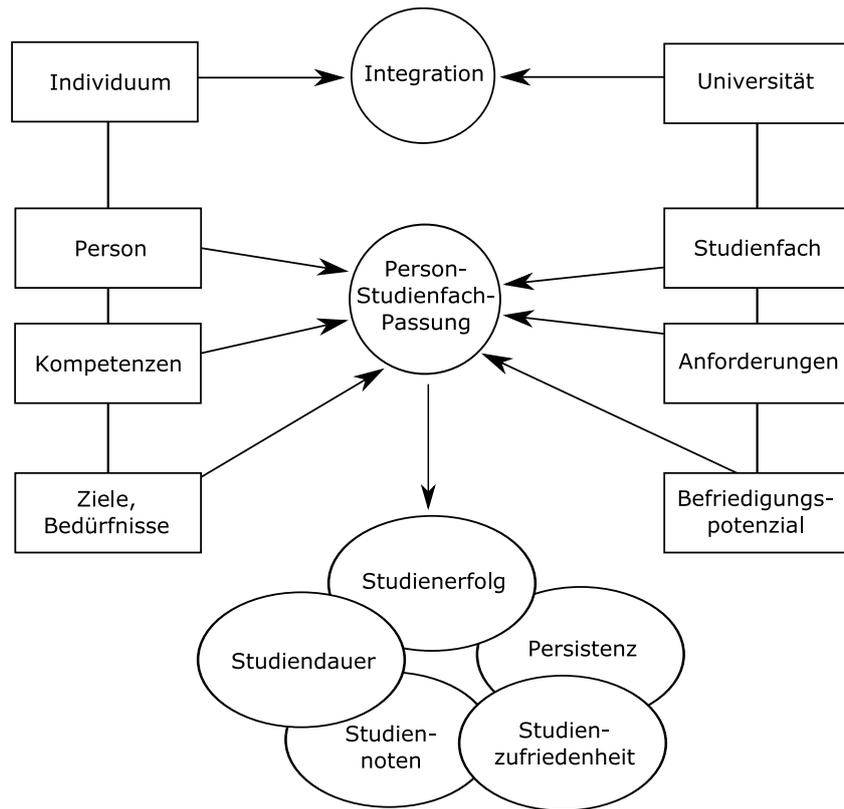


Abbildung 2.1: Person-Studienfach-Passung als Voraussetzung für den Studienerfolg, in Anlehnung an Amelang (1997)

Personen-Environment-Fit, kurz P-E-FIT). Auch wenn häufig in Publikationen die Begriffe Personen-Job-Fit beziehungsweise Personen-Umwelt-Fit als Synonyme verwandt werden, wird hier der Begriff Personen-Umwelt-Fit als Überbegriff verstanden, der sowohl die Passung zu einem bestimmten Beruf (Person-Vocation-Fit), zu konkreten Tätigkeiten an einem Arbeitsplatz (Person-Job-Fit), zu einer Arbeitsgruppe/Team (Person-Group-Fit) und zu einer Organisation (Person-Organization-Fit) (Kristof, 1996) umfasst. Der universitäre Personen-Umwelt-Fit ist im großen Maße abhängig vom gewählten Studiengang, unterscheiden sich doch nicht nur die zu verrichtenden Tätigkeiten in den diversen Studiengängen, sondern auch die Eigenschaften der Dozenten und Kommilitonen und somit die Umwelt.

Die Grafik 2.1 veranschaulicht dieses Zusammenspiel von Person (Individuum) und Umwelt (Universität/Studienfach) und dessen Auswirkungen. Dabei wird sowohl die Passung der Anforderungen und der Fähigkeiten beziehungsweise Kompetenzen (demands-abilities) als auch der Bedürfnisse und des Befriedigungspotenzials (needs-supplies) (Edwards, 1991) berücksichtigt, um ein Verbleiben im Studiengang, eine kurze Studiendauer, hohe Studienzufriedenheit und gute Noten vorhersagen zu können. Untersuchungen zum Studienabbruch zeigen entsprechend auch eine erhöhte Abbruch- und Wechselneigung, wenn die Studenten mit dem Studienangebot nicht zufrieden sind (Heublein, Spangenberg & Sommer, 2003; Spies, Westermann, Heise & Hagen, 1998), also das erwartete Befriedigungspotenzial nicht vorhanden ist. Die Konzeption des Personen-Umwelt-Fits stellt die Grundlage für alle eignungsdiagnostischen Aussagen dar

(Deutsches Institut für Normung, 2002; Schuler, 1996), somit auch für die Studienberatung und Studierendenauswahl.

2.2.2 Die allgemeine und fachspezifische Studierfähigkeit

Obwohl der Begriff der *Studieneignung* in zahlreichen Publikationen erwähnt wird, sind bis dato keine umfassenden Definitionen vorgelegt worden (Konegen-Grenier, 2001). Grundsätzlich gilt, dass eine Person für eine bestimmte Tätigkeit – im hier betrachteten Fall das „Studieren“ – geeignet ist, „wenn sie über diejenigen Merkmale verfügt, die Voraussetzung für die jeweils geforderte berufliche Leistungshöhe sind“ (Deutsches Institut für Normung, 2002, S. 20). Solche Merkmale setzen sich zusammen „aus vorhandenen Interessen, Eigenschaften, Kenntnissen, körperlichen Gegebenheiten und der Fähigkeit, spezifische Anforderungen innerhalb einer bestimmten Zeitspanne erlernen zu können“ (S. 20). Überträgt man diese Definition auf die Studieneignung, so umfasst diese die „Gesamtheit der Fähigkeiten, Kenntnisse, Einstellungen und sonstigen psychischen Merkmale, die wichtig sind für das Durchlaufen [...] eines Studiums“ (Rahn et al., 1977, zitiert nach Blum, 1980, S. 17). Demnach beinhaltet der Begriff *Studieneignung* sowohl Bestandteile der kognitiven und der nicht-kognitiven Merkmalsbereiche. Klar getrennt werden sollten der Begriff *Studierfähigkeit* als kognitive Voraussetzung eines Studiums und der Begriff *Studieneignung*, der – wie auch die oben beschriebene Definition von *Eignung* – alle erfolgsrelevanten Merkmale umfasst. Dieser Einschränkung entsprechend messen Studierfähigkeitstests diejenigen Merkmale im kognitiven Bereich, „die als Voraussetzungen einer erfolgreichen Bewältigung der Anforderungen eines wissenschaftlichen Studiums gelten können“ (Rahn, 1976, S. 11). Um den wesentlichen Einfluss des Studieninteresses auf die Studienfachwahl und den Studienerfolg (Schiefele, Krapp & Schreyer, 1993; Wild, Krapp & Winteler, 1992) zu unterstreichen, wird der Begriff der Studienneigung in Abgrenzung zur Fähigkeitskomponente verwandt.

Heldmann (1984), Kazemzadeh, Minks und Nigmann (1987) und Konegen-Grenier (2001) haben – insbesondere durch Umfragen unter Hochschuldozenten – für den deutschen Sprach- und Bildungsraum Bestimmungsstücke der allgemeinen Studieneignung herausgearbeitet. Kazemzadeh, Minks und Nigmann (1987) unterscheiden *drei Dimensionen der Studieneignung*: *materiale* (fachliche Kenntnisse), *formale* (kognitive Kompetenz, kommunikative Fähigkeiten, Grundlagen wissenschaftlicher Methoden, Kenntnis und Anwendung wissenschaftsbezogener Arbeitstechniken) und *personale Dimensionen* (Lern- und Leistungsbereitschaft, Selbstständigkeit und Motivation, Ausdauer und Belastbarkeit, intellektuelle Neugierde). Die materiale Komponente lässt sich als Vorwissen unter anderem durch die Schulbildung, aber natürlich auch noch später beeinflussen. Die anderen beiden Bereiche gelten als überdauernde Eigenschaften.

Einige Jahre später befragte Konegen-Grenier (2001) wieder Dozenten zur allgemeinen Studieneignung. Zusammenfassend zeigten sich dabei die folgenden Ergebnisse:

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 (S. 168).

Heldmann (1984, S. 132–186) führt fünf Bereiche der Studieneignung mit insgesamt 28 Unterfaktoren an, die im Folgenden genannt werden. Für eine präzisere Beschreibung muss an dieser Stelle aus Platzgründen auf die Originalveröffentlichung verwiesen werden.

- **Ausbildungsbereitschaft**

- Ausbildungsinteresse (Aufmerksamkeit bei den Lernveranstaltungen; Kundgebung des Interesses durch sachliche Fragen; pünktliche Anwesenheit; Einhaltung von vereinbarten Terminen, die für die Ausbildung von Bedeutung sind)
- Interesse der Studienanfänger an persönlicher Studienberatung
- Lernbereitschaft und Leistungsbereitschaft
- Freude am Studium

- **Vorhandensein elementarer Voraussetzungen für wissenschaftliches Arbeiten**

- Vertrautheit mit elementaren Arbeitstechniken (wie Benutzen von Handbüchern, Lexika, Registern, Zitieren, Exzerpieren, Referieren)
- Technik der schriftlichen Darstellung (z.B. Orthografie, Interpunktion)
- Ausdrucksvermögen (sich sprachlich angemessen und differenziert in Wort und Schrift zu äußern)
- Arbeitsqualität (Sorgfalt und Genauigkeit beim Arbeiten sowie mündlichem und schriftlichem Darstellen)
- Präsenz des Wissens (Abrufen von verarbeiteten Kenntnissen)
- Beobachtungsfähigkeit
- Lerntechniken und Problemlösestrategien
- Elementare Kenntnisse der Methoden der Geistes-, Sozial- und Naturwissenschaften (wissenschaftliche Propädeutik)

- **Formen geistigen Tätigseins**

- Allgemeines Denkvermögen
- Abstraktionsvermögen

- Differenzierungsvermögen (Fakten und Meinungen als solche erkennen und entsprechend in der Argumentation einsetzen)
- Auffassungsgabe (Erfassen von Zusammenhängen, Erfassen und Übertragen von Beziehungen)
- Anschauliches Vorstellungsvermögen
- Schöpferische Fähigkeiten (Phantasie und Kreativität, Originalität, Dynamik, Innovationsbereitschaft)
- Intellektuelle Neugier
- Weite des Problembewusstseins (geistige Aufgeschlossenheit, fachübergreifende Interessen, kulturelle Offenheit)
- **Ausprägung der Persönlichkeit**
 - Ichstärke (Selbstvertrauen, Gelassenheit, Risikobereitschaft, Selbstdisziplin, Willensstärke etc.)
 - Ausdauer und Belastbarkeit
 - Psychomotorische Fertigkeit (sportpraktisches Können, Bewegungsfähigkeit)
 - Selbständigkeit und Motivation
 - Urteilsfähigkeit (Skepsis gegenüber eigenen und fremden Aussagen)
- **Interesse und Engagement**
 - Kontaktfähigkeit
 - Interessen außerhalb des Studiums (Theater, Sport, Hobby etc.)
 - Engagierte Haltung (altruistische Einsatzbereitschaft, ethische Berufsauffassung)

Dieses Ergebnis der Anforderungsanalysen deutet auf die besondere Bedeutung nicht-kognitiver Komponenten der Studieneignung hin. Begriffe, die mit dem Interesse, mit der Leistungs- und Lernmotivation, aber auch mit der Selbständigkeit und dem Organisationstalent zusammenhängen, werden hier als wichtig erachtet.

Diese Auflistungen beziehen sich lediglich auf die allgemeine Studieneignung, die klar von den spezifischen Anforderungen bestimmter Studienfächer zu trennen ist. In zahlreichen Anforderungsanalysen (zum Beispiel Blum, 1980; Christian, 1980; Hitpass, 1976; Zimmerhofer, 2003) zeigten sich zwar allgemeine Anforderungen, die für alle Studienfächer gelten, aber auch fachspezifische Besonderheiten und damit eine unterschiedliche Gewichtung mathematischer, räumlicher und sprachlicher Studieninhalte.

Entsprechend differenzierte Anforderungen – selbst im scheinbar homogenen Bereich der mathematischen Vorkenntnisse – konnten im Rahmen der TIMS-Studie (Third International Mathematics and Science Study, Beaton et al., 1996, Mullis et al., 1998) aufgedeckt werden: Dabei wurden Hochschuldozenten aus acht unterschiedlichen Studienfächern um eine Einschätzung der Wichtigkeit unterschiedlicher Mathematikaufgaben aus der TIMSS-Testbatterie gebeten.

Dabei zeigten sich – je nach Studienfach und damit je nach verwandter Mathematik – gravierende Unterschiede in der Einschätzung (Baumert, Bos & Lehmann, 2000). Ferner konnten auf Basis dieser Umfrage Studienfächer zu Studienfeldern zusammengefasst werden, die zumindest im Bereich der Mathematik scheinbar ähnliche Anforderungen haben. Diese Befunde illustrieren, dass zumindest in der subjektiven Sicht der Hochschuldozenten die Definition der fachlichen Voraussetzungen für ein erfolgreiches Studium erheblich zwischen den Studienfächern variieren kann, und zwar selbst dann, wenn es sich lediglich um ein einziges schulisches Fachgebiet, wie hier die Mathematik, handelt.

Um die Studieneignung eines bestimmten Studienfaches beschreiben zu können, reicht es daher nicht, lediglich auf die allgemeinen Eignungskomponenten zurückzugreifen. Vielmehr muss die allgemeine Studierfähigkeit um weitere fachspezifische Komponenten erweitert werden. Anforderungsanalysen in Form von Workshops mit Fachvertretern lassen dabei die Dimension der Umwelt fachspezifisch beschreiben (Deidesheimer Kreis, 1997). Um den Unterschieden der einzelnen Studienfächer Rechnung zu tragen, aber auch die Gemeinsamkeiten dieser abbilden zu können, lassen sich Studienfelder definieren, die aus mehreren, bezüglich ihrer Anforderungen ähnlichen Studienfächern bestehen (Trost & Haase, 2005): In diesem so genannten Blütenmodell zeigen zwei Fächer wie zum Beispiel Sozialwissenschaften und Informatik nur wenige Überlappungen bezüglich verschiedenster Anforderungsdimensionen, Fächer wie Informatik und Mathematik hingegen größere. Eine grafische Darstellung der Konzeption als Blütenmodell lässt sich der Abbildung 2.2 entnehmen.

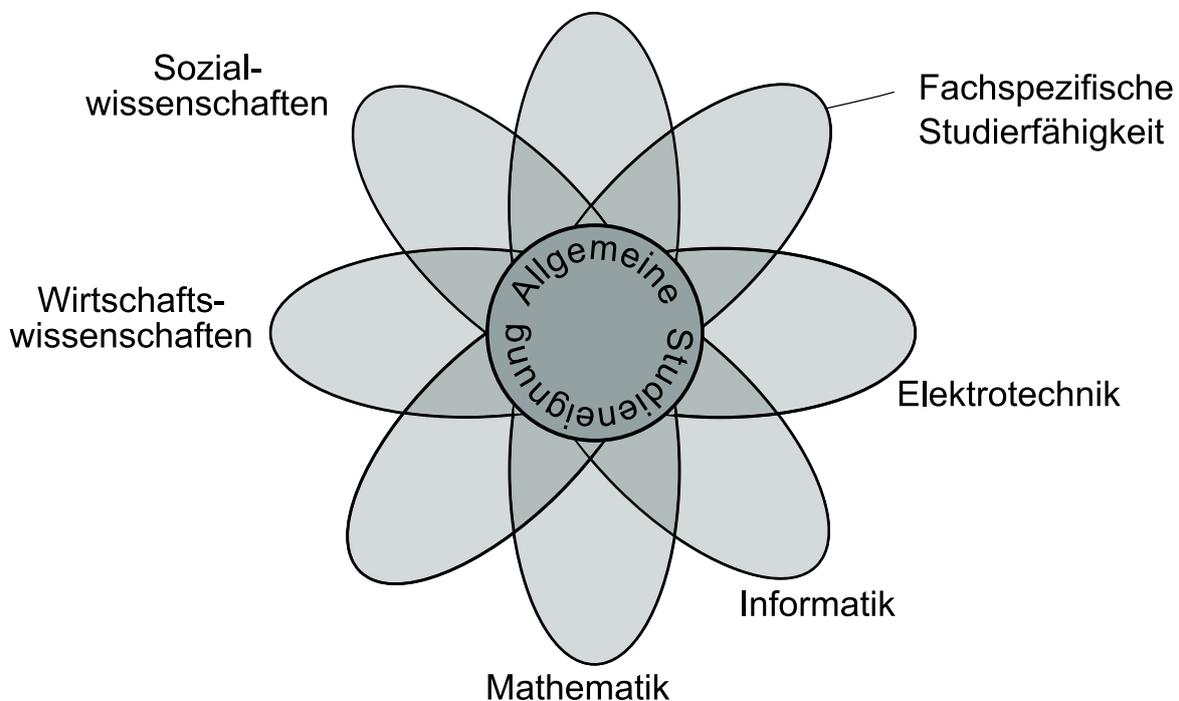


Abbildung 2.2: Darstellung allgemeiner und fachspezifischer Studierfähigkeit (modifiziert nach Trost, 2003)

2.3 Fremdselektion durch die Hochschulen

Seit Jahren wurde immer wieder seitens der Hochschulen und somit auch der Hochschulrektorenkonferenz versucht, den autonomen Handlungsspielraum der ausbildenden Institutionen zu vergrößern und diesen insbesondere mehr Einflussmöglichkeiten bei der Auswahl geeigneter Studieninteressierter zu geben. Statt der allgemeinen Studierfähigkeit, die auf der Basis der Abiturdurchschnittsnote gemessen wurde, sollte verstärkt die fachspezifische Eignung im Mittelpunkt stehen (Amelang, 1997; Deidesheimer Kreis, 1997; Rindermann, 2005; Trost, 1975; Trost & Haase, 2005; Wissenschaftsrat, 2004). Aber auch die Unterschiede der Hochschulen beziehungsweise der Institute sollten im Rahmen der Verfahren differenzierter abgebildet werden können als dies lange Zeit der Fall war. So sollte zumindest Teilen der bereits erwähnten Probleme bezüglich der Qualität, Quantität und insbesondere der niedrigen Erfolgsquoten begegnet werden können. Verbunden war damit aber ebenfalls die Idee, über eine stärker von den Hochschulen bestimmte Studierendenauswahl in der Zulassung passendere Studieninteressierte zu finden, aber auch eine stärkere Profilbildung der deutschen Hochschullandschaft und somit auch eine Elitisierung bestimmter Hochschulen zu erreichen.

Die langjährigen Auseinandersetzungen zur Hochschulautonomie führten zur 7. Novelle des Hochschulrahmengesetzes (Bundesministerium für Bildung und Forschung, 2004a), die im Sommer 2004 mit Zustimmung des Bundesrates vom Bundestag verabschiedet wurde. Damit wurde ein neues Vorgehen für die zentralen Numerus-clausus-Fächer (im bundesweiten Verfahren sind dies die Diplomstudiengänge Biologie, Medizin, Pharmazie, Psychologie, Tiermedizin, Zahnmedizin) festgeschrieben: Nach Abzug der Vorabquoten (für Nicht-EU-Studienbewerber, Härtefälle etc.) werden ein Fünftel der Studienplätze an die Abiturbesten (nach Bundesländern getrennt betrachtet), ein weiteres Fünftel nach der Wartezeit und die verbleibenden 60 Prozent von den Hochschulen selbst vergeben. Durch diese Gesetzesnovelle werden zugleich die Hochschulquote² deutlich erhöht, die Auswahlkriterien erweitert und zusätzlich noch ein Verfahren umgesetzt, in dem die Ergebnisse der Hochschulauswahlverfahren nicht durch Nachrückverfahren der ZVS revidiert werden.

Hochschulen dürfen nach §32 des Hochschulrahmengesetzes (Bundesministerium für Bildung und Forschung, 2004a) Studienplätze vergeben...

- a) nach dem Grad der Qualifikation nach §27 (Abiturnote),
- b) nach gewichteten Einzelnoten der Qualifikation nach §27, die über die fachspezifische Eignung Auskunft geben,
- c) nach dem Ergebnis eines fachspezifischen Studierfähigkeitstests,
- d) nach der Art einer Berufsausbildung oder Berufstätigkeit,

² Die Hochschulquote ist der Anteil der Studienplätze, die eine Hochschule nach eigenen Kriterien vergibt beziehungsweise vergeben lässt.

- e) nach dem Ergebnis eines von der Hochschule durchzuführenden Gesprächs mit den Bewerberinnen und Bewerbern, das Aufschluss über die Motivation der Bewerberin oder des Bewerbers und über die Identifikation mit dem gewählten Studium und dem angestrebten Beruf geben sowie zur Vermeidung von Fehlvorstellungen über die Anforderungen des Studiums dienen soll und
- f) auf Grund einer Verbindung von Maßstäben nach den Buchstaben a bis e (S. 2298).

Übereinstimmend mit den Darlegungen im Abschnitt 2.2.2 hat der Gesetzgeber damit ausdrücklich nur die Verwendung von fachspezifischen Studierfähigkeitstests für die Studierendenauswahl im Hochschulauswahlverfahren (Bundesministerium für Bildung und Forschung, 2005) erlaubt. Der Einsatz eines allgemeinen Studierfähigkeitstests oder gar eines allgemeinen Intelligenztests ist somit nicht für dieses Einsatzgebiet zulässig.

Studienfeldspezifische Eignungstests unterscheiden sich in einigen Aspekten von generellen Fähigkeitstests: Erstens sind die Verfahren anhand von Anforderungsanalysen speziell auf das Studienfach zugeschnitten. Ferner wird eine besondere Kontextualisierung durch die Item-Inhalte angestrebt, um somit eine besondere augenscheinliche Inhaltsnähe zum Studienfach zu erreichen. Auf Basis herkömmlicher Intelligenzmodelle wird zudem eine spezifische Gewichtung relevanter Intelligenzdimensionen (vgl. dazu zum Beispiel Carroll, 1993), aber auch die Erstellung besonderer Normen für die jeweilige Zielgruppe (Trost, 1975; Trost, 2003) angestrebt.

Im Abschnitt 3.3.1 werden die Prädiktionsleistungen unterschiedlicher Verfahren, darunter eben auch die der studienfeldspezifischen Studierfähigkeitstests, miteinander verglichen.

Bei der Konstruktion eines Auswahlprozesses muss zusätzlich beachtet werden, dass der Abiturnote bei der Studienplatzvergabe „ein maßgeblicher Einfluss gegeben wird“ (Bundesministerium für Bildung und Forschung, 2004a, S. 2299). Je nach Umsetzung in die jeweiligen Landesrechte können davon abweichende Vorgaben (zum Beispiel durch eine festgeschriebene Mindestanzahl unterschiedlicher Kriterien, siehe Landesrecht Baden-Württemberg; Gesetzblatt für Baden-Württemberg, 2005) definiert sein. Die Vorgaben sind außerdem nur für die oben genannten zentralen NC-Fächer gegeben, die Vorgaben für weitere Fächer lehnen sich aber – wiederum je nach Bundesland – eng an.

2.4 Selbstselektion durch den Bewerber

Eine aufwändige Studierendenauswahl – wie in Abschnitt 2.3 beschrieben – ist immer dann besonders effizient und nützlich, wenn sich möglichst viele Personen auf wenige Studienplätze bewerben, also eine geringe Selektionsquote vorliegt. Ein modernes Studierendenmarketing, das das Profil einer gesamten Universität oder eines einzelnen Fachbereichs verdeutlicht und

die Vorteile der Wahl unterstreicht, unterstützt somit indirekt auch die Auswahlverfahren, die richtigen Bewerber zu finden. Anders als Hochschulen in den USA, die schon vor Jahrzehnten das Studierendenmarketing einführten und bis heute professionalisiert haben, wird das Werben um die richtigen Studenten in Deutschland erst halbherzig betrieben, allerdings mit einer insgesamt steigenden Tendenz.

Das Ziel, zwischen möglichst vielen Bewerbern auswählen zu können, ist allerdings auch nur ein Ansatzpunkt von zwei möglichen, eine hohe Effizienz der Fremdselektion zu erreichen: Neben der Steigerung der reinen Bewerberanzahl ist es sinnvoll, prozentual viele geeignete und wenige ungeeignete Bewerber (entspricht einer hohen Basisrate, siehe Taylor & Russell, 1939, Wiggins, 1973) am Auswahlverfahren teilnehmen zu lassen. Gerade Personen mit einem passenden Eignungsprofil müssen ermuntert werden, sich zu bewerben und somit am Verfahren teilzunehmen. Sie müssen durch eine umfassende Beratungsleistung in ihrer Entscheidung bestärkt werden, indem ihre besondere Passung und die damit verbundenen Erfolgsaussichten kommuniziert werden. Personen mit einer geringeren Passung müssen hingegen alternative Vorschläge unterbreitet bekommen. Da der Studienbewerber in diesen Fällen selbst – allerdings wohl unterstützt – entscheidet, ob er sich für die Aufnahme des Studiums als geeignet erachtet, wird dieser Schritt in Abgrenzung zur *Fremdselektion* (dort führt die Hochschule oder die ZVS die Auswahl durch) *Selbstselektion* genannt. Eine Unterstützung der Selbstselektion ist also auch dann wichtig, wenn effiziente eignungsdiagnostische Entscheidungen bei der Studierendenauswahl durch die Hochschulen zu treffen sind (Wiggins, 1973). Beide Verfahren greifen somit im optimalen Fall ineinander.

Ist eine solche Fremdselektion aber (zum Beispiel aufgrund hoher Kosten) nicht sinnvoll oder rechtlich nicht erlaubt (zum Beispiel, weil weniger Bewerber als Studienplätze vorhanden sind), ist eine umfassende Studienberatung und damit eine Unterstützung der Selbstselektion noch wichtiger: Je informierter Personen vor ihrer Studienfachwahl sind, desto eher sind diese – wenn sie überhaupt wollen – in der Lage, die Wahrscheinlichkeit für eine drohende Unzufriedenheit mit den Inhalten oder gar einen Studienabbruch etc. zu antizipieren. Zu selten sind aber diese Rahmenbedingungen eines informierten Studienbewerbers gegeben (Bergmann, Brandstätter & Eder, 1994; Heublein & Sommer, 2002): Heine und Willich (2006) konnten in einer Befragungstudie an Schulen aufzeigen, dass knapp die Hälfte der befragten Schüler sechs Monate vor dem Verlassen der Schule dringend Beratung benötigt, um sich Klarheit über die persönliche Eignung zu verschaffen. Jeder Dritte wünscht sich ferner Hilfestellung beim Konkretisieren der persönlichen Interessen.

Aufgrund der schlechten Informationslage der deutschen Schulabgänger sind nur die wenigsten Personen in der Lage, Studienfächer auszuwählen, die zu dem eigenen Eignungsprofil passen (zur Personen-Umwelt-Passung siehe Abschnitt 2.1). Der Wissenschaftsrat kommt nach umfangreicher Analyse daher zum Schluss: „Über die Hälfte der Studienabbrecher insgesamt

und knapp zwei Drittel der Studienabbrecher in den ersten beiden Hochschulsesemestern [geben] als entscheidende Gründe für ihren Studienabbruch solche Motive an, deren Wirksamkeit durch Beratung und Information über Eignung und Studiengestaltung wenigstens in Grenzen beeinflussbar wäre“ (Wissenschaftsrat, 2004, S. 21). Daraus abgeleitet, regt der Wissenschaftsrat in seinen Empfehlungen an, insbesondere Qualität und Reflexionsniveau der Studienentscheidung von Studierwilligen dadurch zu verbessern, „dass Studienberatung und Studieninformation erheblich professionalisiert sowie Studien- und Berufsberatung weit mehr als bisher zusammengeführt werden“ (S. 34). Vergleicht man die Berufs- und Studienberatung in Deutschland mit anderen OECD-Ländern, so fällt auf, dass vergleichsweise wenige Fertigkeiten zur eigenen Entscheidungsfindung (Organisation for Economic Co-operation and Development, 2002; siehe Abschnitt 2.1 zur schwierigen Entscheidungsfindung bei der Studienfachwahl) und zu viele reine Informationen über den Beruf vermittelt werden. Drei von vier Studienberechtigten (74 %) haben sich bei Arbeitsagenturen oder Berufsinformationszentren (BIZ) über Ausbildungs- und Studienmöglichkeiten informiert – wobei lediglich jeder dritter Nutzer (31 %) die so erhaltenen Informationen als ertragreich einstuft. Drei von fünf Studienberechtigten (58 %) nehmen persönliche Studien- und Berufsberatungsangebote wahr; lediglich gut jeder dritte Nutzer bewertet den daraus gezogenen Ertrag positiv.

2.5 Internetgestützte Berufs- und Studienberatung

Die Bedeutung von Berufsberatung im Allgemeinen konnte in US-amerikanischen Metaanalysen (Oliver & Spokane, 1988; Whiston, Sexton & Lasoff, 1998) eindrücklich nachgewiesen werden: So zeigt sich ein positiver Einfluss von Beratungsleistungen zum Beispiel auf die Anzahl der in Betracht gezogenen Alternativen, auf die antizipierte Zufriedenheit mit der Wahl, auf das Wissen über sich und über Karrierewege im Allgemeinen. In einer Analyse bezüglich des österreichischen Hochschulsystems konnte ein positiver Beratungseffekt auf Prüfungsnoten, auf die Selbsteinschätzung der Studienleistungen, Studienzufriedenheit und auf die selbst eingeschätzte Stabilität der Studienwahl aufgezeigt werden (Brandstätter, Grillich & Farthofer, 2002).

Insbesondere das Internet wird heute von nahezu allen Studienberechtigten (93 %) zur Informationsbeschaffung verwendet und trägt durch die vielseitigen Informationen zur Komplexitätsreduktion bei der Entscheidungsfindung bei (Heine & Willich, 2006). 71 % der Nutzer dieser Quelle konnten einen hohen Ertrag daraus ziehen. Gerade bei der Selbstselektion (siehe Abschnitt 2.4) bietet sich die Nutzung des Mediums an, lassen sich doch mit dieser Technik nicht nur alleine klassische Informationen über das Studienfach (zum Beispiel eine Beschreibung der Studienfächer, der möglichen Studienschwerpunkte und der späteren Arbeitsmarktchancen) beziehungsweise über das Berufsfeld (zum Beispiel eine Beschreibung der Arbeitstätigkeiten, der

möglichen Arbeitgeber, der Arbeitsmarktchancen und der durchschnittlichen Verdienste) darbieten. Anders als gedruckte Informationen können interaktive Systeme helfen, die Analyse der Passung zwischen Bewerber und den Anforderungen (siehe Abschnitt 2.2.1) voranzutreiben. Für die umfassende Karriereberatung hat sich der Einsatz von computer- und internetgestützten Systemen in den USA schon seit Jahrzehnten bewährt, zeigen sie doch ähnliche positive Effekte wie bei klassischen Beratungen mit einem menschlichen Berater, so genannte *Face-to-Face-Beratungen* (Whiston, Sexton & Lasoff, 1998). Anders als in den USA bestehen in Deutschland bis heute nur wenige Beratungsangebote als so genannte *guidance at a distance* (Organisation for Economic Co-operation and Development, 2002). Diese Form der Beratung erfolgt nicht räumlich gebunden als Gespräch eines Interessenten mit einem Berufsberater. Vielmehr findet der Kontakt dabei zumeist frei zugänglich über das Internet oder aber über andere Medien, die keine Anwesenheit erfordern, statt.

Viele Vorbehalte in der Benutzung des Internets für psychologische Fragestellungen sind in den letzten Jahren ausgeräumt worden (als Übersichtsartikel siehe Gosling, Vazire, Srivastava & John, 2004), so dass sich nicht nur immer mehr Forschungsfragen mit dem Internet selbst beschäftigen (beispielsweise Kraut et al., 1998; Renner, Marcus, Machilek & Schütz, 2005), sondern das Internet verstärkt als komfortables Medium zur Beschleunigung des wissenschaftlichen Austausches, aber auch der wissenschaftlichen Überprüfung von Hypothesen jeglicher Art verstanden wird. Dabei ist insbesondere die Nutzung des Internets für die Verbreitung von Testverfahren und Fragebogen in den letzten Jahren komfortabler geworden und wird entsprechend auch intensiver genutzt – sowohl für die psychologische Forschung (z.B. Bandilla, 1999; Barak, 1999; Birnbaum, 2000; Buchanan & Smith, 1999) als auch für die Personalauswahl von Industrieunternehmen (Bartram, 2000; Hünninghausen & Brümmer, 2002; Wild, De la Fontaine & Schafstetter, 2001). Trotz der großen Unterschiede der Medien wird heute von der allgemeinen Generalisierbarkeit der über das Medium Internet gefundenen Effekte ausgegangen (Kraut et al., 2004; Naglieri et al., 2004). Entscheidend dabei ist, dass zu vermutende Medieneffekte für sorgfältig konstruierte, lokale, PC-gestützte kognitive Tests (Mead & Drasgow, 1993), für Persönlichkeitstests (Richman, Kiesler, Weisband & Drasgow, 1999) und für internetgestützte Testungen (Salgado & Moscoso, 2003) weitgehend auszuschließen sind. Stichprobenverzerrungen sanken in den letzten Jahren in Folge des enormen Wachstums von Internetanschlüssen (deutschlandweit siehe Forschungsgruppe Wahlen Online, 2006; europaweit siehe Nielsen-NetRatings, 2004; weltweit siehe Nielsen-NetRatings, 2003). Im direkten Vergleich zu den ansonsten häufig verwendeten Stichproben herkömmlicher Versuchsdurchführungen – zum Beispiel Studenten – ist gar eine geringere Verzerrung anzunehmen (Gosling, Vazire, Srivastava & John, 2004).

Gerade für den Bereich der Berufs- und insbesondere der Studienberatung werden im Folgenden die Vor-, aber auch die Nachteile der internetgestützten *guidance at a distance* im Vergleich zur klassischen Face-to-Face-Beratung erörtert.

Folgende Vorteile der internetgestützten *guidance at a distance* lassen sich formulieren:

- *Asynchronität*: Anders als bei der üblichen Face-to-Face-Beratung kann die webbasierte Studienberatung rund um die Uhr an sieben Tagen in der Woche stattfinden. Aufwändige Terminabsprachen oder Wartezeiten entfallen für den Ratsuchenden (Batinic & Bosnjak, 2000).
- *Alokalität*: Der Zugang zur Beratung ist nicht räumlich-institutionell beschränkt (Batinic & Bosnjak, 2000). Durch die Verwendung des Internets können weltweit Testungen durchgeführt werden. Eine Anreise zum Test- und Beratungsort und damit auch teilweise hohe Reisekosten entfallen, was insbesondere für Studieninteressierte aus anderen Ländern relevant ist.
- *Höhere Anonymität*: Die herkömmliche Studienberatung erfolgt meist in einer Face-to-Face-Situation, wobei das Problem der *social-desirability* (soziale Erwünschtheit), aber auch der *self-disclosure* (geringe Bereitschaft, Informationen über die eigene Person zu geben) auftreten können. Webgestützte Beratungssysteme finden hingegen meist anonym, also ohne Bekanntgabe von Namen und postalischer Adresse statt. Dadurch vermindert sich der Anreiz, die eigene Einschätzung zu verfälschen und die eigenen Eigenschaften „im besseren Licht“ darzustellen (*impression management*) (Richman, Kiesler, Weisband & Drasgow, 1999; Rosenfeld, Booth-Kewley, Edwards & Thomas, 1996). Schon in den Instruktionen kann somit glaubhaft dargestellt werden, dass die Teilnehmer sich nur selbst belügen, wenn sie sich besser darstellen würden, als sie wirklich sind.
- *Hohe Objektivität*: Die computerisierte Berufsberatung ist aufgrund des Automatisierungsgrades komplett unabhängig von den Fachkenntnissen und der Interpretationsleistung eines menschlichen Berufsberaters. Bei gleichen Rahmenbedingungen werden aufgrund der standardisierten Vorgabe und der regelgeleiteten Interpretation den Ratsuchenden immer exakt dieselben Empfehlungen gegeben. Die Testsituation selbst ist hingegen nur dann annähernd standardisiert, wenn die Ratsuchenden die entsprechenden Empfehlungen bezüglich des Testumfeldes umsetzen.
- *Hohe Ökonomie*: Durch den hohen Automatisierungsgrad sind die Kosten pro Beratungsleistung bei einer größeren Anzahl an Teilnehmern gering. Anders als die Durchführung von Face-To-Face-Beratungen bestehen die Kosten meist nur aus fixen Ausgaben, so dass durch eine Erhöhung der Teilnehmerzahlen die Pro-Kopf-Kosten drastisch sinken. Aufgrund der guten Skalierbarkeit der technischen Voraussetzungen (Webserver, Netzwerktechnik etc.) kann selbst bei größtem Andrang der Betrieb gewährleistet werden. Bei einer klassischen Beratung müssten in diesem Fall lange Wartezeiten in Kauf genommen werden.
- *Vorteile für die Evaluation*: Das Internet ist ein beliebtes Informationsmedium für Schüler (Heine & Willich, 2006), wodurch sich gerade für die Evaluation hohe Stichprobengrößen ergeben. Ferner stehen die einzelnen Datensätze der Beratungsleistungen nicht dezentral,

sondern leicht zugänglich in einer elektronischen Datenbank zur Verfügung (Bartram, 2003). Eine Evaluation wird somit komfortabler und kostengünstiger.

- *Hohe Marketingwirkung*: Aufgrund der besonderen Zielgruppennähe haben moderne webbasierte Verfahren einen imagebildenden Charakter. Hochschulen können somit auch ihr Profil positiv verändern, wenn sie ein solches Verfahren einsetzen (Zimmerhofer, Heukamp & Hornke, 2006).

Die genannten Vorteile einer internetgestützten Beratung gehen mit einigen kritischen Aspekten einher, die im Folgenden genannt werden:

- *Fehlende menschliche Interaktionspartner*: Ein wesentlicher Bestandteil der computerisierten Karriereberatung ist die automatisierte Testinterpretation (kurz CBTI). Nach Sampson, Purgar und Shy (2003) integriert CBTI „theory, practitioner judgment, and empirical evidence, as one source of data, to help practitioners and test takers better understand the meaning of test scores for the purpose of gaining insight and making decisions“ (S. 25). Anders als ein Gespräch mit einem menschlichen Studienberater können automatisierte Beratungstools allerdings nicht auf alle Besonderheiten des Teilnehmers eingehen. Es steht kein Ansprechpartner unmittelbar zur Verfügung, der beim Verstehen der Rückmeldung und deren Interpretation unterstützen, Missverständnisse ausräumen und weitere Schritte zusammen mit dem Interessenten planen kann.
- *Gefahr einer Überbewertung*: Wie alle Beratungsleistungen – also auch die herkömmliche Face-to-Face-Beratung – kann eine webbasierte Berufs- oder Studienberatung nur eine Facette in einem gesamten Prozess zur optimalen Entscheidung darstellen. Wichtig ist es auch nach einem erfolgten Ergebnis, weitere Informationen zu berücksichtigen, andere Menschen anzusprechen, über mögliche weitere Studienalternativen nachzudenken etc. Durch die leichte Verfügbarkeit, die hohe Anonymität der Online-Verfahren und ein generelles Vertrauen in die Technik kann die Gefahr leicht entstehen, dass die webbasierte Facette zur einzigen Informationsquelle wird.
- *Fehlende Standardisierung der Testumgebung*: Ein entscheidender Vorteil, nämlich die Alokalität, also die räumliche Unabhängigkeit, wird schnell zu einem essentiellen Nachteil: Anders als bei Face-to-Face-Beratungen müssen die Testkandidaten, die eine Beratung im Internet aufsuchen, selbst eine ruhige Arbeitsatmosphäre einrichten, damit die Ergebnisse auch eine gute Vorhersagekraft für den Berufs- oder Studienerfolg haben. Ebenso problematisch ist die unerlaubte Verwendung von Hilfsmitteln, die Personen „nur mal eben“ zur Unterstützung heranziehen und dann die Prognosefähigkeit des gesamten Verfahrens unbewusst stark einschränken. Wesentlich ist es daher, in der Instruktion auf Faktoren hinzuweisen, die die Validität der Beratungsempfehlung mindern.
- *Eingeschränkte Steuerfunktion*: Webbasierte Tools zur Karriereberatung unterstützen lediglich die Selbstselektion durch den Bewerber. Die Fremdselektion gerade bei populären Studienfächern können sie jedoch nicht ersetzen. Wesentlicher Unterschied zwischen

den beiden Methoden ist nämlich, dass die Ergebnisse des Beratungstools lediglich eine Empfehlung darstellen, bei einer Fremdselektion sind diese hingegen verpflichtend. Der Studieninteressent entscheidet im Rahmen der Selbstselektion weiterhin eigenständig, ob er ein bestimmtes Studium aufnimmt oder nicht. Entsprechend ist die Steuerfunktion der Beratungssysteme, also der Einfluss des Beratungsergebnisses auf die schlussendliche Entscheidung, stark eingeschränkt. Entscheidend für die Steuerfunktion ist die Augenscheinvalidität und die Akzeptanz des Verfahrens (Hell & Schuler, 2005).

Für derartige Studienberatungstools lassen sich eine Vielzahl unterschiedlicher Bezeichnungen in Publikationen, aber auch im Internet finden. Sowohl „Career-Assessment“ (Barak, 2003), „Career-Guidance“ (Offer & Sampson, 1999) oder „Self-Assessment“ (Zimmerhofer & Hornke, 2005b) umschreiben ähnliche Ansätze zur Berufs- und Studienberatung. Sehr unterschiedlich ist hingegen der Aufbau bei gleicher Intention (Oliver & Zack, 1999).

Einige Beispiele, die die Vielseitigkeit der webbasierten Berufs- oder Studienberatung eindrücklich aufzeigen:

- *The Career Key* (<http://www.careerkey.org>) oder *Career Builder* (<http://www.careerpath.com>) bieten auf der Basis der Persönlichkeitstheorie nach Holland (1997) eine Auswahl passender Berufe. Damit unterstützt das System Ratsuchende dabei, mögliche Berufsalternativen kennen zu lernen, die zu den eigenen Interessen passen. Leistungstests werden nicht als Ergänzung angeboten.
- Das *O*NET* (Occupational Information Network, <http://www.onetcenter.org>) ist eine umfassende Quelle für berufsbezogene Informationen: Primär stellt das staatliche O*NET Konsortium aus den USA eine Datenbank dar, die seit Ende 2005 für 580 Berufe die Anforderungen in den Bereichen Wissen, Fertigkeiten, Fähigkeiten und Eigenschaften – also die Ergebnisse zahlreicher Anforderungsanalysen – aufzählt (Peterson et al., 2001). Eher für Wissenschaftler oder Berufsberater konzipiert, ist das System für Ratsuchende wenig geeignet. Eine Prüfung auf die lediglich aufgezählten Kompetenzen etc. der einzelnen Berufe geschieht nicht.
- Das Hochschul-E-Assessment-Project der Technischen Universität Hamburg-Harburg und der Universität Hamburg *HEAP* (<http://www.studienkompass-hh.de>) entwickelte speziell für die Fächer Psychologie, Wirtschaftswissenschaften, Informatik und Maschinenbau/Produktion webbasierte Beratungsverfahren. Dabei werden sowohl Kenntnisse über die Studieninhalte (z.B. über multimediale Inhalte) vermittelt als auch in Fragebogen und Tests die Eignung und Neigung für diese Fächer getrennt abgeprüft. Der Schwerpunkt der derzeitigen Version liegt aber auf einer multimedialen Einführung in die genannten Fächer, nicht auf einer Eignungsprüfung.
- Die *LMU München* (<http://www.pms.ifi.lmu.de/eignungstest>) ermöglicht Ratsuchenden Aussagen über die eigene Eignung für ein Informatikstudium. Dabei werden 12 Aufgaben

(aus den Bereichen Logik, Mathematik, Englisch und Deutsch) dargeboten und die Anzahl der richtigen Lösungen zurückgemeldet. Eine statistische Evaluation des Testmaterials ist nach Aussage der Betreiber nicht durchgeführt worden.

- Die *Eligo GmbH* entwickelte für die Ruhr-Universität Bochum das Beratungstool *Borakel* (<http://www.ruhr-uni-bochum.de/borakel>). Anhand von Fragebogen und Leistungstests wird der Ratsuchende sowohl bei der Planung des weiteren Berufswegs als auch bei der Wahl des richtigen Studienfachs unterstützt. Auch bietet das System durch einen multi-medialen Anteil Einblicke in das Universitätsleben.
- Das Forschungsprojekt *Career Counseling for Teachers* (<http://www.cct-germany.de>) ermöglicht, selbst zu erkunden, ob der Lehrerberuf zur eigenen Interessenstruktur, aber auch zur Persönlichkeit passt. Ferner hilft ein nicht-kognitiver Fragebogen bei der Auswahl der Unterrichtsfächer.

Der bereits oben genannte Begriff *Self-Assessment* soll im Folgenden präziser definiert werden, umschreibt er doch sehr eingängig die Intention entsprechender Beratungssysteme:

Der Begriff *Self-Assessment* wird weltweit zahlreich eingesetzt (Dunn, McEntarffer & Halonen, 2004; Kotter, 1978), indizierte doch alleine Google 30,3 Millionen Seiten (Google, 2006), die diesen Begriff enthielten. Die Kombination aus „*self*“ (Selbst) und „*assessment*“ (Beurteilung, Einschätzung) impliziert im Kontext der Karriereberatung eine Möglichkeit, die eigene Eignung für verschiedene Berufe oder Studienfächer zu erkennen – ganz ohne dass eine bestimmte Fremdselektion damit verbunden ist. Loacker (2000, zitiert nach Dunn, McEntarffer & Halonen, 2004, S. 173) sieht in einem Self-Assessment einen selbst gesteuerten Prozess, „of observing, analyzing, and judging one’s performance against established standards while identifying ways to improve it“. Bei der Konzeption der Self-Assessments steht somit im Vordergrund, die Reflexionsfähigkeit der Teilnehmer zu trainieren und Verzerrungen der eigenen Eignung zu begegnen.

Zimmerhofer und Hornke (2005b) definieren den Begriff Self-Assessment im Kontext der Studienfachwahl wie folgt (S. 273):

- Ein Self-Assessment ist ein psychometrisches Testverfahren, das eigenständig durchgeführt wird, teilweise auch selbst ausgewertet wird.
- Es stellt einen Baustein zur Unterstützung einer willentlichen und wohlüberlegten Entscheidung (z.B. für die Studienfachwahl) dar, indem mehr Informationen über die eigene Eignung, Neigung und die Anforderungen zur Verfügung stehen.
- Somit dient es auch dazu, Verzerrungen in der Wahrnehmung der eigenen Stärken und Schwächen aufzudecken und mit objektiven Daten abzugleichen.
- Die Teilnahme an dem Beratungsinstrument geschieht freiwillig.
- Die Nicht-Teilnahme hat keine – über den Wegfall des Nutzens hinausgehenden – negativen Konsequenzen.

- Eine allgemein verständliche Rückmeldung ist integraler Bestandteil des Systems.
- Die Rückmeldungen stehen zur persönlichen Weiterentwicklung zur Verfügung und sind für Dritte (z.B. Arbeitgeber oder Hochschule) nicht, nur anonymisiert oder auf Initiative des Ratsuchenden zugänglich.

Der Schwerpunkt dieser Konzeption liegt in der objektiven, aber auch in der subjektiven Verbesserung der Entscheidungsfindung und somit der Selbstselektion. Erreicht werden soll dies durch eine freiwillige Eignungsüberprüfung und weniger durch die Gabe von expliziten Beschreibungen zum Studium oder Beruf. Anders als bei der Fremdselektion kann durch diese Self-Assessments nicht nur die Studierfähigkeit der Studieninteressierten näher betrachtet werden, also eine Prüfung auf kognitive Prädiktoren erfolgen. Ergänzt werden lässt sich die Eignungsprüfung zum Beispiel durch Verfahren zur Lern- und Leistungsmotivation, zum Interesse, zur Handlungsorientierung etc. (siehe Abschnitt 3.4). Gerade diese zeigen sich als wesentliche Prädiktoren, die aber aus Gründen der leichten Verfälschbarkeit nicht in Verfahren zur Auswahl aufgenommen werden können.

2.6 Zusammenfassung

Im vorliegenden Kapitel wurden Ansätze aufgezeigt, die unter anderem dazu führen sollen, dass nicht nur mehr Personen ein Studium beginnen, sondern vor allem mehr Personen ein zu ihren Stärken und Schwächen passendes Studium aufnehmen. Bei einem Entschluss zu einem Studienfach dürfen nicht nur die Arbeitsmarktchancen, der Hochschulort, das Ansehen des Faches etc. berücksichtigt werden. Vielmehr muss die individuelle Eignung und Neigung der Personen im Mittelpunkt stehen, also gerade die erfolgskritischen Faktoren, die der Einzelne nur schwer ohne Hilfe abschätzen kann.

Dabei wurden die Selbst- und Fremdselektion als eine Einheit vorgestellt: Die Selbstselektion durch Beratung bildet die erste Phase, die Auswahl die zweite Phase, deren Einsatz allerdings nur dann opportun ist, wenn mehr Bewerber als Studienplätze vorhanden sind. Durch die internetbasierte Studienberatung kann in der ersten Phase erreicht werden, dass die Personen besser über das eigene Eignungsprofil und über die Anforderungen verschiedenster Studienfächer informiert sind. Auf dieser Basis können sie selbstständig eine fundiertere Wahl treffen, aber auch weitere Gespräche mit Fachstudienberatern oder Beratern der zentralen Studienberatung suchen.

Die Selbstselektion wird bei einer verbesserten Beratung niemals die Fremdselektion, also die Studierendenauswahl, substituieren können. Diese wesentliche Einschränkung rührt aus dem rein empfehlenden Charakter: Empfehlungen können, müssen aber von den Interessierten nicht berücksichtigt werden. Dies wird auch dann für besonders populäre Fächer gelten, wenn die

allgemeine Zahl der Hochschulberechtigten abgesunken ist (Kultusministerkonferenz der Bundesrepublik Deutschland, 2003). Auf jeden Fall kann aber mit einem Beratungssystem erreicht werden, dass sich mehr passende Personen (und somit eine höhere Basisrate) für die Auswahlverfahren anmelden, was wiederum die Kosten des folgenden Auswahltrittes senkt und zugleich die Treffergenauigkeit erhöht.

Gerade für den Bereich der Selbstselektion ist wesentlich, dass diese Informationen rund um die Eignung und Neigung genau dort vorliegen, wo sie am einfachsten abgerufen werden können: zu Hause und 24 Stunden täglich. Genau in der Erreichbarkeit liegt zurzeit noch ein wesentliches Grundproblem der herkömmlichen Studien- und Berufsberatung, die keine oder nur wenige Angebote als *guidance at a distance* anbieten (Organisation for Economic Co-operation and Development, 2002).

Kapitel 3

Kriterien, Verfahren und Prognostizierbarkeit des Studienerfolgs

Im folgenden Kapitel werden zuerst Kriterien für den Erfolg in einem universitären Studiengang samt ihrer Charakteristika erläutert. Diese Übersicht stellt die Grundlage dar, um im nächsten Schritt geeignete Prädiktoren für die einzelnen Kriterien des Studienerfolgs zu benennen (siehe Abschnitt 3.2 bis Abschnitt 3.4.6).

3.1 Studienerfolg – Definition möglicher Kriterien

Es werden im Folgenden drei unterschiedliche Formen des Studienerfolgs voneinander unterschieden: der *persönliche*, der *institutionelle* und der *marktliche* Studienerfolg (Stebler, 2000). Von einem *persönlichen Studienerfolg* lässt sich immer dann sprechen, wenn das Individuum selbst die Erfahrungen an der Universität positiv bewertet. Da sich Lebensziele von Menschen aber nicht gleichen, somit auch die daraus abgeleiteten Studienziele (Viebahn, 1990) variieren, kann die Bewertung des eigenen Studienverlaufs durchaus von objektiven Kennwerten differieren. Entscheidend ist die subjektive Einschätzung des Betroffenen. Die meisten Studierenden sehen das Studium als Möglichkeit zum *Erwerb formaler Qualifikations- und Tätigkeitsnachweise*. Das Studium wird von diesen als ein Instrument angesehen, später einen bestimmten Beruf etc. ausüben zu dürfen, der mit bestimmten Studienabschlüssen verbunden ist. So ist ein Medizinstudium und ein bestandenes Staatsexamen die Voraussetzung für eine Arbeit als Arzt. Das wichtigste anvisierte Ziel ist entsprechend ein erfolgreicher Studienabschluss. Ein Studienabbruch würde hingegen einem subjektiven Misserfolg gleichkommen. Personen, die hingegen *Qualifikationen für bestimmte Tätigkeitsbereiche* durch das Studium anvisieren, können auch nach einem Studienabbruch einen subjektiven Erfolg verspüren – immerhin haben sie vielleicht die erwünschten Qualifikationen bereits erworben. Formale Nachweise stehen dabei nicht im Mittelpunkt. Ähnliches gilt für Personen, die eine *persönliche Entwicklung* durch ein Studium

anstreben. Somit können Personen auch dann von einem Studienerfolg sprechen, wenn sie zwar keinen Abschluss erreicht haben, das Studium aber dennoch für die Persönlichkeitsentwicklung förderlich war oder bestimmte Fertigkeiten erlernt wurden. Diese Studienziele werden zumindest teilweise auch bei einem Studienabbruch erfüllt, wenn auch nicht in dem Umfang, den ein Studienabschluss vermuten lässt. Gerade bestimmte Arbeitstechniken, die in einem Studium erlernt wurden, können auch ohne einen erfolgten Abschluss beruflich erfolgreich eingesetzt werden (Heublein, Spangenberg & Sommer, 2003).

Diese bis hierhin genannten individuellen Ziele sind klar von den gesellschaftlichen Zielen eines Studiums zu trennen: Von einem *institutionellen Studienerfolg* (Stebler, 2000), also einem Studienerfolg aus der Hochschulperspektive betrachtet, lässt sich immer dann zweifelsfrei sprechen, wenn ein Studium in möglichst kurzer Zeit mit einer möglichst guten Note absolviert worden ist. Im Mittelpunkt steht dabei, dass die geringen Ressourcen der Hochschulen (Räume, Lehrpersonal etc.) möglichst effizient genutzt werden.

Die Betrachtung dieses *institutionellen Studienerfolgs* endet dort, wo die Bewährung der Person auf dem Arbeitsmarkt beginnt. Ob die entsprechende Person aufgrund ihrer Ausbildung dort erfolgreich ist, beeinflusst den *marktlichen Studienerfolg*. Selbst Personen, die sowohl persönlich als auch institutionell erfolgreich waren, können am Arbeitsmarkt große Schwierigkeiten haben, einen ausbildungsäquivalenten Arbeitsplatz zu erhalten. Zu viele Studenten werden „am Markt vorbei ausgebildet“ (Koch & Mohr, 2006), haben also ein Qualifikationsprofil, das vom Arbeitsmarkt nicht in diesem Umfang benötigt wird. Trifft dieses auf viele Personen zu, sinkt die durchschnittliche Rentabilität der Bildungsausgaben für die Volkswirtschaft. Die knappen Ressourcen werden dann in Bereiche investiert, die sich nicht entsprechend für die Volkswirtschaft nutzen lassen. Dieses Beispiel verdeutlicht auch, dass die einzelnen Ziele häufig nicht gleichzeitig erreicht werden können. So kann ein bestimmter Studienabschluss vielleicht auf dem Arbeitsmarkt nicht gefragt sein, trotzdem aber zu einer Persönlichkeits- und auch Qualifikationsentwicklung geführt haben.

3.1.1 Studienabschluss versus Studienabbruch

Im Folgenden sollen Möglichkeiten zur Operationalisierung des Studienerfolgs erarbeitet werden.

Um den Studienerfolg zu quantifizieren, lässt sich die dichotome Variable „erfolgreicher Studienabschluss“ versus „Studienabbruch“ auswerten. Damit steht ein simples Erfolgs- bzw. Misserfolgskriterium zur Verfügung, das eng mit dem Studienziel *Erwerb formaler Qualifikations- und Tätigkeitsnachweise* einhergeht (Amelang, 1978). Da die Zeiten bis zum Abbruch nicht selten mehr als 12 Semester betragen (nach Heublein, Spangenberg & Sommer, 2003 etwa in 25 % der Fälle), ist eine umfassende Analyse dieser Kriterien jedoch erst sehr spät möglich.

Eine wesentliche weitere Schwierigkeit ergibt sich dadurch, dass der Misserfolg – bei strenger Definition des Begriffs Studienabbruch – zu Lebzeiten nicht zweifelsfrei festgestellt werden kann (siehe auch 1.4.1), da eine Wiederaufnahme des Studiums auch nach einer längeren Pause (z.B. durch einen Erziehungsurlaub oder Krankheit) immer noch möglich ist.

3.1.2 Studiennoten

Ebenso können Studienabschluss- beziehungsweise Zwischenprüfungsnoten als Kriterien verwendet werden. Diese einfach zur Verfügung stehenden Kriterien werden erwartungsgemäß am häufigsten in Analysen betrachtet (zum Beispiel Baron-Boldt, 1989; Kuncel, Hezlett & Ones, 2001; Kuncel, Hezlett & Ones, 2004; Robbins et al., 2004), auch wenn eigentlich diese Noten nur Abstufungen des Erfolges beziehungsweise der grundsätzlichen Eignung darstellen (Amelang, 1978). Strenggenommen stellen sie keine Negativkriterien dar. Aus Gründen der Praktikabilität werden in zahlreichen Studien nicht Studienabschlussnoten, sondern Noten aus den ersten vier Semestern oder auch aus den ersten zwei Semestern hinzu gezogen. Auch wenn Zwischennoten und Studienabschlussnoten miteinander hoch korrelieren¹, wird mit der Verwendung von Studiennoten aus den Anfangssemestern der Zusammenhang zum Beispiel zwischen Schulnoten und dem entsprechend definierten Studienerfolg überschätzt. Erwartungsgemäß korrelieren nämlich Schulleistungen höher mit den Studiennoten aus den ersten Semestern als mit Studienabschlussnoten (Brandstätter & Farthofer, 2003b).

Ein zentraler Kritikpunkt, der gegen eine Verwendung von Studiennoten spricht, ist die mangelnde psychometrische Qualität. Amelang (1978) geht davon aus, dass die Reliabilität und Validität von Studiennoten im direkten Vergleich zu Schulnoten geringer ausfällt. Gerade unterschiedliche Bewertungsmaßstäbe zwischen Lehrstühlen, Fächern und ganzen Hochschulen führen zudem dazu, dass sich Studiennoten kaum vergleichen lassen. In den meisten Studienfächern zeigt sich ferner eine sehr eingeschränkte Varianz in den vergebenen Noten (Wissenschaftsrat, 2003), die die Korrelationen zu den Prädiktoren verringert. Fraglich ist ebenfalls, was Studiennoten überhaupt messen: Neben einem motivationalem Konglomerat scheinen Prüfungen in den meisten Fällen lediglich Wissen abzufragen (Rosas Díaz, 1990).

Studiennoten stellen – trotz dieser Unzulänglichkeiten – ein wesentliches Kriterium für die Erfassung des Studienerfolgs dar. Sie sind nämlich – aufgrund der diagnostisch-filternden Funktion (Gold & Souvignier, 1997) – besonders praxisrelevant: So zeigt sich, dass gerade Personen, die kurz nach dem Examen einen Arbeitsplatz gefunden haben, prozentual häufiger ebenfalls eine gute Examensnote aufweisen (Abele, Andrä & Schute, 1999). Entsprechend geht von den Studiennoten ein direkter Einfluss auf die Chancen am Arbeitsmarkt (Rindermann & Oubaid,

¹ Es werden mittlere bis hohe Korrelationen zwischen Zwischenprüfungen und Examensnoten von $r = 0,38$ (Ingenieurwissenschaften), $r = 0,42$ (Medizin) und $r = 0,54$ (Wirtschaftswissenschaften) berichtet (Gold & Souvignier, 2005).

1999), auf das Einstiegsgehalt und somit auf die gesamte Karriereentwicklung aus. Infolgedessen ist das Kriterium für den Studienerfolg gleichzeitig ein Prädiktor für den Erfolg im Berufsleben.

3.1.3 Studiendauer

Die durchschnittliche Dauer eines grundständigen Studiums wird – ähnlich wie die allgemeinen Erfolgsquoten – in Deutschland diskutiert und gilt als ein wesentlicher Leistungsindikator eines Hochschulsystems.

Welche Einflüsse die höhere Verweildauer deutscher Studierender auf die volkswirtschaftlichen Kosten für einen Studienplatz hat, zeigen eindrücklich die internationalen Vergleiche der OECD (Organisation für Wirtschaftliche Zusammenarbeit und Entwicklung, 2004): Jedes zusätzliche Studienjahr schiebt den Start ins Berufsleben heraus und lässt somit die individuellen Ertragsraten des universitären Abschlusses sinken. Aufgrund der hohen jährlichen Kosten für den Staat sinkt auch die Ertragsrate für die Volkswirtschaft mit der Länge der Ausbildung. Die durchschnittliche Studiendauer im Tertiärbereich A beträgt in Deutschland 6,5 Jahre. Nur für Griechenland (Dauer 8,1 Jahre) hat die OECD (Organisation für Wirtschaftliche Zusammenarbeit und Entwicklung, 2004) ein höheres Mittel berechnet. Im OECD-Durchschnitt verlassen Studenten schon nach 4,7 Jahren ihre Hochschule. Über diese gesamte Verweildauer eines Studenten kostet das Studium im Durchschnitt von 19 OECD-Ländern 43 000 \$ pro Student. In Deutschland hingegen kostet ein Studium – auch aufgrund der längeren Studiendauer – durchschnittlich 55 426 \$. Betrachtet man nämlich die Ausgaben pro Jahr (Tertiärbereich ohne Ausgaben der Forschung und Entwicklung) pro Student, so liegt Deutschland mit 6 370 \$ weit unter dem OECD-Mittel (10 724 \$).

Eine wesentliche Frage ist natürlich, wer für die langen Studienzeiten verantwortlich ist – eher das Individuum oder eher die Institution beziehungsweise das Bildungssystem. Durch stark variierende Studienordnungen, die Studienbedingungen, den Stoffumfang, die Wiederholungsmöglichkeiten nicht bestandener Prüfungen und lange Korrekturphasen lassen sich 60 % bis 70 % der Varianz in der Studiendauer erklären (Reissert, 1991). Demnach klärt der Personenfaktor nur 30 % bis 40 % der Varianz auf. Dieser besondere Effekt der Umwelt zeigt sich auch, wenn man die durchschnittliche Studiendauer verschiedener Hochschulen vergleicht: Große Unterschiede lassen sich zwischen den einzelnen Hochschulen (Wissenschaftsrat, 2001) nachweisen, obwohl davon auszugehen ist, dass sich die Personeneigenschaften – zum Beispiel die Leistungsbereitschaft – der Studenten an den unterschiedlichen Hochschulorten nicht signifikant unterscheiden. Wenn bestimmte Faktoren der Institution entscheidend sind, dürfte eine Prädiktion der Studiendauer durch eine personenbezogene Eignungsdiagnostik nur schwerlich

gelingen (Höppel & Moser, 1993). Giesen und Gold verdeutlichen die besondere Bedeutung der Umweltfaktoren in dem folgenden Vergleich von Studenten mit Marathonläufern:

Wenn ein Marathonläufer wegen schlechter Markierung von der Strecke abkommt, nutzen Appelle an die Motivation und unterstützende Maßnahmen wenig. (S. 97)

Aber natürlich lassen sich auch starke individuelle Einflussgrößen auf die Studiendauer finden: Eines der in den letzten Jahren häufig beachteten Konzepte ist die so genannte *academic procrastination*: Vom lateinischen *procrastinare*² abgeleitet, werden unter diesen Begriff dysfunktionales Verschieben, Aufschieben, Vermeidungsverhalten und Unpünktlichkeit subsumiert (Helmke & Schrader, 2000). Die Auswirkungen dieses Phänomens sind vielseitig: Als *trait* oder *state* verstanden, gehen damit das Aufschieben kleiner, lästiger Haushaltspflichten, die Verschleppung von Prüfungsvorbereitungen bis hin zum Aufschieben elementarer Lebensentscheidungen einher. Personen mit einer hohen *academic procrastination* als *trait* investieren entsprechend weniger Lernzeit, zeigen geringeres Studieninteresse, verschieben häufiger Prüfungen und studieren entsprechend länger als ihre Kommilitonen (Helmke & Schrader, 2000). Operationalisieren lässt sich *procrastination* als abweichende Anzahl bestandener Prüfungen zu bestimmten Zeitpunkten, zum Beispiel nach vier Semestern. Anders als mit dem Kriterium Studiendauer lässt sich eine Auswertung bereits nach einigen Semestern durchführen.

3.1.4 Studienzufriedenheit

Neben den Studiennoten ist die Studienzufriedenheit das am häufigsten verwendete Kriterium für den globalen Studienerfolg, wenn es auch aufgrund seiner Breite ganz unterschiedlich operationalisiert wird. So kann die Umsetzung von der „Integration in die Gruppe der Kommilitonen bis zur Selbsteinschätzung guter Studienfortschritte reichen“ (Giesen et al., 1981, S. 71). Verstanden wird die Studienzufriedenheit meist in Analogie zu dem Konzept der Arbeitszufriedenheit, also als positive beziehungsweise negative Bewertung des eigenen Studiums (Spies, Westermann, Heise & Hagen, 1998). Die jeweiligen Bewertungen entsprechen ferner der „emotionale[n] Reaktion auf die Arbeit, [der] Meinung über die Arbeit und [der] Bereitschaft, sich in der Arbeit zu engagieren“ (Nerdinger, 2001, S. 351). Dabei hängt die Zufriedenheit ähnlich wie schon die Studienleistungen und Persistenz von der Passung zwischen Person und Umwelt ab (Assouline & Meir, 1987; Spies, Westermann, Heise & Hagen, 1998; Tranberg, Slane & Ekeberg, 1993; Tsabari, Tziner & Meir, 2005).

Westermann, Heise, Spies und Trautwein (1996) führen zehn Hauptmerkmale der Studienzufriedenheit an, deren Gewichtung von Student zu Student differiert:

- *Studienbedingungen* (allgemeine Bedingungen und Umstände an der Hochschule)

² wörtlich: etwas auf einen anderen Tag verlegen oder verschieben.

- *Studieninhalte* (Nützlichkeit der Studieninhalte und Interesse für diese)
- *Lehrverhalten der Lehrenden* (empfundene Kompetenz der Dozenten und Sympathie für diese)
- *Benotung* (wahrgenommene gerechte Verteilung der Noten und Zufriedenheit mit diesen)
- *Berufliche Relevanz* (Zukunftsansichten mit einem Studienabschluss)
- *Anerkennung* (Meinung anderer Personen über das Studienfach)
- *Kontakte zu Lehrenden* (Grad der Kontakte zu Dozenten)
- *Unterstützende Kontakte zu Kommilitonen* (soziales Klima unter Kommilitonen)
- *Hochschulverwaltung und -politik* (Kundenorientierung der Hochschulverwaltung)
- *Randbedingungen des Studiums* (Einfluss persönlicher Bedingungen auf die allgemeine Zufriedenheit)

Eine Verwendung der Studienzufriedenheit als Erfolgskriterium und somit auch als Kriterium für die Evaluation von Studiaauswahlverfahren ist umstritten, hängt doch die Wahrnehmung der Studienzufriedenheit besonders von den individuell anvisierten Studienzielen ab (siehe Abschnitt 3.1), aber auch von den erhaltenen Noten. So kann angenommen werden, dass gerade die Zufriedenheit mit der eigenen Benotung andere Aspekte der Zufriedenheit überstrahlt.

3.1.5 Zusammenfassung

Die vorigen Ausführungen fassten unterschiedlichste Kriterien des Studienerfolgs zusammen. Verdeutlicht wurde dabei, dass es nicht *den* Studienerfolg geben kann. Was letztendlich als Studienerfolg definiert wird, hängt in einem nicht unbeträchtlichen Maße vom Betrachter und – im Falle von empirischen Überprüfungen – auch von der Praktikabilität der Erfassung ab. So wird der Student selbst eine andere Perspektive auf den Gegenstandsbereich haben als dessen Eltern, als die Hochschulverwaltung oder Bildungsforscher. Klarer wurde auch, warum Studiennoten das häufigste Erfolgskriterium darstellen. Gerade Noten der ersten vier Semester stehen vergleichsweise leicht für Analysen zur Verfügung und sind wiederum sehr gute Prädiktoren für den weiteren Studienverlauf. Als weiteres zentrales Kriterium wurde die Anzahl der Prüfungsleistung zu bestimmten Zeitpunkten erwähnt. Dies lässt sich als eine direkte Operationalisierung der *academic procrastination* verstehen und zudem – anders als die Studiendauer – nicht erst nach mehreren Jahren präzise messen.

Die Unterschiedlichkeit der genannten Kriterien verdeutlicht, dass auch die Prädiktoren entsprechend differenziert betrachtet werden müssen. Anzunehmen ist, dass die Prognose unterschiedlicher Kriterien (zum Beispiel Studiennoten oder Studiendauer) durch unterschiedliche Kriterien erfolgreich vorhergesagt werden kann. Im Folgenden werden daher verschiedene Prädiktoren vorgestellt, die in der Literatur genannt und in der Praxis, sei es in der testgestützten Studienberatung oder bei der Studierendenauswahl, eingesetzt werden.

3.2 Hochschulreifezeugnis als Prädiktor

Im Folgenden werden die wichtigsten Prädiktoren für die bereits dargestellten Studienerfolgsriterien (siehe Abschnitt 3.1) genannt. Dabei sollen insbesondere die Reliabilität und Validität, aber auch die Effizienz des Einsatzes betrachtet werden. Begonnen wird mit der genaueren Betrachtung der Vorhersage des Studienerfolgs durch Schulnoten, insbesondere durch die Abiturgesamtnote.

Die besondere Bedeutung des Abiturs als Hochschulreifezeugnis ist unbestritten, besitzt doch ganz automatisch jede Person (siehe §27 des Hochschulrahmengesetzes HRG, Bahro & Berlin, 2003), die das Abitur erworben hat, ein Recht auf Zugang zum gewünschten Hochschulstudium (Kultusministerkonferenz der Bundesrepublik Deutschland, 2000). Die prinzipiell uneingeschränkte Wahlfreiheit des Studienfaches beruht auf einer Bildungsidee, nach der die Schulen – trotz der heterogenen Konzeption heutiger Oberstufen (mit unterschiedlicher zeitlicher Dauer, unterschiedlichen Schwerpunkten beziehungsweise Abwahlmöglichkeiten) – die grundsätzliche Eignung in allen akademischen Fächern hinreichend sichern. Eingeschränkt wird diese Freiheit hingegen bei Studiengängen, in denen mehr Bewerber als Studienplätze gezählt werden. Die Hochschulzulassung in diesen zulassungsbeschränkten Studienfächern wird meist über einen zentralen oder hochschulbezogenen Numerus clausus geregelt, bei der die Abiturdurchschnittsnote als Kriterium zur Zulassung herangezogen wird. Aber auch bei diesen Fächern entscheidet der Numerus clausus nur über das *Wann*, nicht hingegen über das *Ob*. Nach einer Wartezeit kann somit jede Person, die die allgemeine Hochschulreife erlangt hat, das Fach der Wahl studieren.

Diese herausragende Bedeutung des Abiturs beziehungsweise der Abiturnoten gilt auch nach der 7. Novelle des Hochschulrahmengesetzes, die im Sommer 2004 vom Bundesgesetzgeber verabschiedet wurde. So ist darin festgeschrieben, dass die Allgemeine Hochschulreife mit der entsprechenden Note auch bei hochschuleigenen Auswahlverfahren (HAV) einen „maßgeblichen Einfluss“ haben muss (siehe Abschnitt 2.3). Eine Verwendung der Abiturdurchschnittsnote als Auswahlkriterium besitzt für Hochschulen darüber hinaus einige praktische Vorteile: Die Abiturdurchschnittsnote ist bei allen inländischen Bewerbern an Universitäten stets vorhanden³. Die Berücksichtigung ist einfach und kostengünstig, da keine aufwändigen Testverfahren, Interviews etc. berücksichtigt werden müssen. Entsprechend wird das Abiturzeugnis häufig als „Inbegriff der Hochschulreife und als Garant der faktischen Studierfähigkeit“ (Wolter, 1989, S. 62) verstanden.

Dabei gibt es seit der Einführung des allgemeinen Hochschulreifezeugnisses Bedenken „über Art, Umfang und Verbindlichkeit des gymnasialen Lehrplans“, so dass auch „ganz generell

³ Bei Bewerbern der Fachhochschulen ist die Abiturdurchschnittsnote aufgrund der diversifizierten Zugangswege nicht immer verfügbar.

Leistung und Aufgabenerfüllung des Gymnasiums als Ort der Studienvorbereitung zur Debatte standen“ (Wolter, 1989, S. 63). Entsprechend ist auch die Prognosefähigkeit der Abiturnoten umstritten. Ingenkamp (1991) kommt hinsichtlich der psychometrischen Güte von Schulnoten zusammenfassend zum vernichtenden Schluss, dass „Zensuren weder objektiv, noch zuverlässig (reliabel) oder gültig (valide) sind“ (S. 12). Diese üblichen Gütekriterien für psychometrische Testverfahren (Lienert & Raatz, 1998) werden im Folgenden zur Analyse von Schulnoten herangezogen: Die Objektivität fordert die Unabhängigkeit des Messvorgangs von der Subjektivität des Messenden (Ingenkamp, 1991). Unterschiedliches Lehrpersonal müsste idealerweise die Leistungen der Schüler mit gleichen Aufgabenstellungen und unter den gleichen Arbeitsbedingungen messen (Durchführungsobjektivität), die Ergebnisse der Schüler in Klassenarbeiten, aber auch im mündlichen Unterricht nach gleichen Kriterien auswerten (Auswertungsobjektivität) und diese Ergebnisse gleich beurteilen (Interpretationsobjektivität). Die Reliabilität beschreibt, inwieweit ein Messinstrument beim Messen desselben Merkmals dasselbe Messergebnis erhält beziehungsweise inwieweit zwei Messinstrumente identische Messergebnisse erzielen. Übertragen auf das Schüler-Lehrer-Verhältnis ist der Lehrer mit dem Messinstrument gleichzusetzen. Wie im Bereich der Personalentscheidungen in Unternehmen, zum Beispiel bei Assessment-Center (Spychalski, Quinones, Gaugler & Pohley, 1997) oder Interview (Trost, 1996), lassen sich Beobachtungsmängel, Erinnerungsfehler oder Urteilstendenzen auch im Schulalltag kaum vermeiden, allenfalls reduzieren. Die Vergabe der jeweiligen Schulnoten hängt also möglicherweise in vielen Fällen nicht nur von den einzelnen Schülerleistungen ab, sondern zusätzlich von der jeweiligen Benotungsgewohnheit (Tendenz zur Milde oder zur Strenge) eines Lehrers und von weiteren Beurteilungsfehlern wie dem Halo-Effekt (Beeinflussung der wahrgenommenen Leistungsfähigkeit durch besonders positiv bewertete Eigenschaften, d.h. gute Schüler erhalten in vielen Bewertungssituationen einen Bonus, schwächere hingegen einen Malus) (Stegmann, 1996). Demzufolge lassen sich Schulnoten, die von unterschiedlichen Lehrern vergeben wurden, nur eingeschränkt vergleichen. Aber auch die Bewertungen eines Lehrers hängen nicht in jedem Fall wirklich von den Leistungen des Schülers ab, sondern eben auch von der Stärke des Klassenverbundes. Helfen würde zweierlei: Informationen über die Verteilung der Schulnoten im Klassenverbund (Modalwert, Median, Mittelwert, Standardabweichung etc.) oder den entsprechenden Rangplatz des Schülers im Leistungsvergleich der Klasse würden unterstützen, Zeugnisse – gerade trotz Noteninflation – besser vergleichen und zugleich Leistungsunterschiede besser aufzeigen zu können.

Um ferner die wichtigen Schulabschlüsse, so insbesondere das Abitur, auch mit Blick auf die nachgewiesenen Lernniveaus vergleichbarer und transparenter zu gestalten, sind alle Bestrebungen in den Bundesländern zu unterstützen, Schulabschlussprüfungen künftig zentral zu organisieren. Somit wird die Fokussierung auf die Inputvariablen – festgelegt durch die Lehrpläne, die Lehrer bei der Unterrichtsgestaltung zu berücksichtigen haben – durch eine Orientierung auf den Output – festgelegt durch die zentralen Abituraufgaben – verschoben. Anzunehmen ist,

dass durch eine einheitliche Prüfung, einheitliche Auswerteregeln und Notenvergabe die Vorhersagbarkeit des Berufs- und Studienerfolgs durch die Schulnoten steigt. Allerdings ist nach der derzeitigen Planung in Nordrhein-Westfalen lediglich die Aufgabenerstellung Bestandteil der zentralen Organisation, die Auswertung und Notenvergabe obliegt – wenn auch mit klarer definierten Kriterien – weiterhin den einzelnen Lehrern. Ein Problem bleibt allerdings auch nach der Einführung des Zentralabiturs: Schülerleistungen können lediglich konfundiert mit der Unterrichts- beziehungsweise Lehrerqualität erhoben werden.

Trotz der eingeschränkten Objektivität und Reliabilität von Einzelnoten deuten sowohl US-amerikanische (Burton & Ramist, 2001; Robbins et al., 2004) als auch deutsche Metaanalysen (Baron-Boldt, 1989; Hell, Trapmann, Weigand, Hirn & Schuler, 2005; Trapmann, Hell, Weigand & Schuler, 2007) auf eine passable Vorhersage des Studienerfolgs durch das Abitur beziehungsweise durch die Abschlussnote der High-School hin. Der Schwerpunkt der Betrachtung wird im Folgenden auf deutsche Studien gelegt, da die Ergebnisse aus den US-amerikanischen Metaanalysen nur sehr begrenzt auf die deutschen Verhältnisse übertragbar sind.

Vor etwa 15 Jahren führte Baron-Boldt (Baron-Boldt, Schuler & Funke, 1988; Baron-Boldt, 1989; Schuler, Funke & Baron-Boldt, 1990) die erste große Metaanalyse zur Prädiktionsleistung der Abiturnote in Deutschland durch. Dabei zeigte sich eine korrigierte Korrelation⁴ der Abiturdurchschnittsnote mit dem Studienerfolg (je nach betrachteter Studie definiert als „bestanden“ versus „nicht-bestanden“, aber auch „gutes Resultat“ versus „schwaches Resultat“) von $r = 0,456$ (95 %-Konfidenzintervall liegt zwischen 0,317 und 0,595). Führt man eine Unterscheidung der Studienfächer ein, so zeigt sich eine Schwankung der Korrelationen zwischen 0,377 (Rechtswissenschaft) und 0,557 (Wirtschaftswissenschaften). Trapmann, Hell, Weigand und Schuler (2007) führten in den letzten Jahren eine ähnliche Metaanalyse mit europäischen Daten durch: Dabei zeigte sich eine korrigierte Korrelation von $r = 0,47$ ⁵ beziehungsweise $r = 0,53$ ⁶.

Bei den Analysen zum Test für medizinische Studiengänge (TMS) zeigte sich, dass die Abiturdurchschnittsnote mit dem schriftlichen Teil der Ärztlichen Vorprüfung mit $r = 0,48$, mit dem mündlichen Teil zu $r = 0,34$ beziehungsweise 0,47 mit der Gesamtnote korrelierten (Trost, Klieme & Nauels, 1997).

US-amerikanische Studien kommen zu ähnlichen Prädiktionsergebnissen, wenn auch das Bildungssystem mit dem deutschen nur sehr eingeschränkt verglichen werden kann. In einer Metaanalyse erzielten Robbins, Lauver, Le, Davis, Langley und Carlstrom (2004) eine korrigierte Korrelation⁷ von $r = 0,413$ ($n = 17\,196$), Burton und Ramist (2001) einen Zusammenhang von $r = 0,49$ ($n = 9\,675$). Das Kriterium beider Studien war der Grade Point Average (GPA),

⁴ Eine Korrektur für die Varianzeinschränkung wurde durchgeführt.

⁵ Eine Korrektur für die Reliabilität des Kriteriums wurde durchgeführt.

⁶ Eine Korrektur für die Reliabilität des Kriteriums und für die Varianzeinschränkung wurden durchgeführt.

⁷ Die Ergebnisse sind korrigiert für den Messfehler im Kriterium, nicht im Prädiktor.

also die Durchschnittsnote des Abschlusszeugnisses am College. Geringere Korrelationen werden berichtet, wenn nicht der Notendurchschnitt des Colleges berücksichtigt wird, sondern das Ergebnis einer Lizenzierung: So zeigt der High School GPA eine unkorrigierte Korrelation von $r = 0,16$ mit den Noten in der Lizenzierung NABPLEX für Pharmazeuten (Kuncel et al., 2005) ($n = 509$), eine höhere korrigierte Korrelation⁸ von maximal $r = 0,540$ hingegen bei der Prädiktion der Zwischennoten am College.

Die meisten bisher dargestellten empirischen Ergebnisse versuchten, die Prädiktionsleistungen von Zeugnisnoten für Studiennoten abzuschätzen. Wie bereits im vorigen Abschnitt 3.1 dargestellt, stellt dies aber nur eines von mehreren möglichen Kriterien dar. Der Studienabbruch, aber auch der Fachwechsel lassen sich im Allgemeinen weniger gut vorhersagen, sind aber genauso praxisrelevant. Die umfangreichste Metaanalyse zu dieser Fragestellung stammt dabei aus den USA und erfasst 80 000 Studienbewerber an 380 Universitäten. Unkorrigiert wurde dabei ein statistischer Zusammenhang zwischen den High-School-Noten und dem Studienabschluss (binär „bestanden“ versus „nicht bestanden“) von $r = 0,290$ errechnet (Burton & Ramist, 2001). Robbins et al. (2004) berichten leicht geringere Korrelationen, nämlich unkorrigiert $r = 0,239$ beziehungsweise korrigiert⁸ $r = 0,240$ ($n = 5\,551$). In einer Längsschnittanalyse ($n = 3\,500$) konnte Weck (1991) einen signifikanten Unterschied (bei einem Signifikanzniveau von 0,05) zwischen Schulnoten in Mathematik, Physik, Deutsch und Geschichte von Fachwechslern und konstant Studierenden aufzeigen. In Englisch, Französisch und Latein gelang dies hingegen nicht.

Selten wurde die Studiendauer als Vorhersagekriterium näher betrachtet: Im Rahmen der ärztlichen Vorprüfung konnte Hinneberg (2003) nachweisen, dass bessere Abiturienten im Mittel die Prüfung eher ablegten. Höppel und Moser (1993) konnten mithilfe von Schulabschlussnoten etwa 8 % der Varianz in der Studiendauer aufklären. Im direkten Vergleich zu der Varianzaufklärung in den Abschlussnoten von 18 % ist diese jedoch gering. Giesen und Gold (1996) differenzierten dieses Ergebnis nach unterschiedlichen Studienfächern: So konnten sie Mittelwertsunterschiede in der Abiturnote zwischen lang studierenden und schnell studierenden Personen der Mathematik und Naturwissenschaften nachweisen. Bei anderen Fächern, insbesondere der Geisteswissenschaften, war eine Prognose der Studiendauer weder durch das Abitur noch durch die Allgemeine Intelligenz möglich.

Neben der Verwendung der Durchschnittsnoten werden immer wieder Versuche unternommen, die Prädiktionsleistung verschiedener Studienerfolgskriterien durch eine Berücksichtigung von Einzelnoten zu optimieren (zum Beispiel Fries, 2002; Gold & Souvignier, 2005; Meier, 2003). Die Grundidee in diesem Vorgehen liegt darin, die fachspezifischen Besonderheiten in die Gewichtung der Einzelnoten einfließen zu lassen und entsprechend auch Noten nicht zu berücksichtigen, die wie Sport oder Musik in den meisten Fällen keine kognitiven Fähigkeiten

⁸ Die Ergebnisse sind korrigiert für den Messfehler im Kriterium, nicht im Prädiktor.

abbilden. In diversen Analysen konnte aufgezeigt werden, dass gerade die Mathematiknote der beste Einzelprädiktor ist (zusammenfassend Trapmann, Hell, Weigand & Schuler, 2007). In der bereits erwähnten Metaanalyse von Hell et al. (2005) zeigt sich eine korrigierte Korrelation⁹ von $r = 0,27$. Ergebnisse aus der eignungsdiagnostischen Forschung unterstreichen die besondere Bedeutung der Mathematik (zum Beispiel Althoff, 1986; Schuler, 1998; Stegmann, 1996). Ähnlich wichtig ist die Beherrschung der eigenen Landessprache (also zum Beispiel Französisch in Frankreich, Deutsch in Deutschland), deren Schulnote mit dem Studienerfolg alleine zu $r = 0,22$ korreliert (Hell, Trapmann, Weigand, Hirn & Schuler, 2005).

Trotz der besonderen Augenscheinvalidität von gewichteten Einzelnoten zeigt sich die Abiturdurchschnittsnote bei den meisten Studien der besonderen Gewichtung von bestimmten Einzelnoten überlegen (Deidesheimer Kreis, 1997; Lissmann, 1977; Rindermann & Oubaid, 1999). Die Abiturdurchschnittsnote integriert – eher als Einzelfachnoten – Allgemeinbildung, eine generelle mentale Fähigkeit, aber eben auch nicht-kognitive Kompetenzen, also Lern- und Leistungsmotivation. Aufgrund der Aggregation über einen langen Zeitraum (meist zwei Jahre), über mehrere Lehrer, über unterschiedliche Schulfächer zeigt sich letztlich eine höhere Reliabilität der Abiturgesamtnote (Schuler, Funke & Baron-Boldt, 1990) und auch eine bessere Validität. Demgemäß können Einzelnoten nur in wenigen Ausnahmen – je nach Studien- und Schulfach – eine inkrementelle Validität über die Abiturdurchschnittsnote aufzeigen (Gold & Souvignier, 2005; Trapmann, Hell, Hirn, Weigand & Schuler, 2005; Trapmann, Hell, Weigand & Schuler, 2007).

Zum Beispiel konnten Moosbrugger, Jonkisz und Fucks (2006) zwar für das Fach Psychologie zeigen, dass sich die Varianzaufklärung durch ein Hinzunehmen der Anzahl erlernter Fremdsprachen und der entsprechend unterrichteten Halbjahre von 15 % auf 29 % steigern ließ. Fraglich ist nur, ob diese zusätzlichen varianzaufklärenden Einzelnoten und weiteren Informationen auch einer Kreuzvalidierung über verschiedene Stichproben und gar Universitäten standhalten würden. Dabei bleibt ferner die Frage, wie vorgegangen werden soll, wenn ein in der Varianzaufklärung zentrales Schulfach – auch bei einer sehr guten Abiturdurchschnittsnote – frühzeitig abgewählt wurde. Ein genereller Ausschluss eines solchen Studienbewerbers ist wahrscheinlich kaum juristisch durchsetzbar und meist auch nicht sinnvoll, sollen doch gerade gute Abiturienten ein Studium aufnehmen. Zu beachten ist dabei, dass die entsprechenden Ergebnisse nur auf Basis von Korrelationsanalysen empirisch aufgefunden wurden. Eine inhaltliche Begründung, warum zum Beispiel die von Moosbrugger, Jonkisz und Fucks (2006) aufgezeigte Anzahl erlernter Fremdsprachen etwa über die Zulassung zu einem Psychologiestudium entscheiden sollte, lässt sich nicht nachvollziehen. Entsprechende Rechtfertigungen – eventuell vor Gericht – sind somit schwer zu führen.

⁹ Die Ergebnisse sind korrigiert für den Messfehler im Kriterium, nicht im Prädiktor.

Das Abitur, operationalisiert als Abiturgesamtnote, zeigte sich in zahlreichen nationalen und internationalen Studien als bester Einzelprädiktor des späteren Studienerfolgs, wobei aber lediglich die allgemeine Studierfähigkeit erfasst wird. Für eine fachspezifische Aussage könnte eine Gewichtung der Schulnoten vorgenommen werden, allerdings mit den bereits erwähnten praktischen Problemen. Für eine fachspezifische Eignungsdiagnostik, gerade auch im Bereich der Studienberatung, müssen daher weitere Prädiktoren gesucht werden, die im Folgenden näher vorgestellt werden.

3.3 Kognitive Verfahren als Prädiktoren

3.3.1 Allgemeine Intelligenz

Kognitive Fähigkeiten beeinflussen alle Aspekte des Lebens – in der Arbeit und in der Freizeit gleichermaßen. Entsprechendes lässt sich auch für den Bereich des Studiums vermuten. Brand (1987) vergleicht – übereinstimmend mit dieser allgemeinen Einschätzung – die Bedeutung der Intelligenz für die Psychologie mit der Bedeutung von Kohlenstoff für die Chemie.

Schon früh definierten Autoren Facetten des Begriffs: So umschreibt Thorndike (1921, zitiert nach Jensen, 1998, S. 47) Intelligenz mit „the power of good responses from the point of view of truth and fact“ oder Terman (1921, zitiert nach Jensen, 1998, S. 47) mit „the power of think abstractly“. Stern betont die Anpassungsfähigkeit an neue Aufgaben und Bedingungen des Lebens (Brocke & Beauducel, 2001), Rohracher (1988) die Fähigkeit, bei neuen Aufgaben bisherige Erfahrungen zu verwerten.

Spearman führte Anfang des 20. Jahrhunderts den Begriff der *General Mental Ability* ein. Er vermutete, dass „all branches of intellectual activity have in common one fundamental function (or group of functions), whereas the remaining or specific elements of the activity seem in every case to be wholly different from that in all the others“ (Spearman, 1904, S. 284). Damit postulierte er einen gemeinsamen Faktor der Intelligenz – kurz *g* –, der durch weitere spezifische Intelligenzfaktoren – kurz *s* – ergänzt wird. Gottfredson (1997, zitiert nach Lubinski, 2004) fasst zahlreiche Definitionen zusammen und entwirft die *General Mental Ability* als „a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience“ (S. 97).

Zahlreiche US-amerikanische und europäische Metaanalysen bewiesen die große Bedeutung der generellen Intelligenz (im Folgenden *GMA* abgekürzt) sowohl für den Wissenserwerb im Beruf, für den Ausbildungserfolg im Unternehmen beziehungsweise für die Leistung im Beruf (Hunter & Hunter, 1984; Salgado et al., 2003; Salgado, Anderson, Moscoso, Bertua & de Fruyt, 2003; Schmidt & Hunter, 2004), für das spätere Einkommen (Judge, Higgins, Thoresen &

Barrick, 1999), aber auch für den Schulerfolg (Amthauer, Brocke, Liepmann & Beauducel, 2001; Huldi, 1997; Kleine & Jäger, 1989; Süß, 2001), für die akademischen Leistungen an Universitäten (Kleine & Jäger, 1989; Kuncel, Hezlett & Ones, 2004; Rosas Díaz, 1990), für die Studiendauer (Giesen & Gold, 1996) und longitudinal gar für einen kompletten Karriereweg (Judge, Higgins, Thoresen & Barrick, 1999) oder für gesundheitliche Probleme (Lubinski & Humphreys, 1997). Dabei zeigen empirische Ergebnisse, dass die Bedeutung der *GMA* gar noch steigt, wenn die Komplexität der Arbeit zunimmt (Gottfredson, 1997).

Erwartungsgemäß klären allgemeine Intelligenztests vergleichsweise viel Varianz in dem Kriterium Studiennoten auf, insbesondere in den Fächern der Naturwissenschaften. So konnte Rosas Díaz (1990) für die Prädiktion der Studiennoten durch die Allgemeine Intelligenz – gemessen mit dem Berliner-Intelligenz-Struktur-Test BIS – in den Naturwissenschaften Korrelationen zwischen $r = 0,20$ und $0,25$ (unkorrigiert) nachweisen. Eine höhere unkorrigierte Korrelation ($r = 0,36$) mit dem Studienerfolg an Fachhochschulen berichtet Kersting (1999). Dabei wurde ebenfalls ein Testverfahren auf Basis des Berliner-Intelligenzstruktur-Modells verwandt. Kuncel (2004) konnte ähnliche Ergebnisse für US-Hochschulen zeigen: So korrelierte Millers Analogietest, ein Test der allgemeinen Intelligenz, mit den Studiennoten in seiner Metaanalyse zu $r = 0,39^{10}$ ($n = 11\,368$) und mit der Studiendauer zu $r = 0,35$ ($n = 1\,700$).

3.3.2 Spezifische Komponenten der Intelligenz

Die allgemeine Intelligenz ist ein ökonomischer, valider und zugleich universell nutzbarer Prädiktor – für Trainingserfolg und Arbeitsleistung gleichermaßen (Ree & Carretta, 2002; Schmidt, 2002; Schmidt & Hunter, 2004). Durchschnittlich klärt aber der allgemeine Intelligenzfaktor g nur knapp mehr als die Hälfte der gesamten Testvarianz auf (Carroll, 1993); viel Raum also für weitere Faktoren beziehungsweise Faktorebenen, wie sie sowohl im Modell von Carroll (1993) als auch im Modell von Jäger (1973) zu finden sind. Immer wieder wird versucht, die Prädiktionsleistungen im Vergleich zu einem Test der allgemeinen Intelligenz durch Tests zu verbessern, die die spezifische Intelligenz s abdecken. Für den Bereich der Hochschulzulassung werden statt allgemeiner Intelligenztests allgemeine oder fachspezifische Studierfähigkeitstests entwickelt. Rechtlich wird gar in Deutschland ausschließlich der Einsatz letzterer Verfahren vorgesehen (siehe Abschnitt 2.3).

Gerade in den USA existieren langjährige Erfahrungen mit fachspezifischen Studierfähigkeitstests (Bennett, 2005; Hubin, 1988; Paterson, Schneidler & Williamson, 1938). Das High-School-Zeugnis (*high-school GPA*) entscheidet an den meisten US-Colleges zusammen mit dem Testergebnis des SAT (ehemals scholastic aptitude test) des Colleges Board (entwickelt und durchgeführt durch den Educational Testing Service) oder des ACT (ehemals American College Testing

¹⁰ Die Ergebnisse sind doppelt minderungskorrigiert mit Berücksichtigung der Varianzeinschränkung.

Program) über den Zugang (Harackiewicz, Barron, Tauer & Elliot, 2002). Weitere Tests, so zum Beispiel der GRE (Graduate Record Examinations) oder der GMAT (Graduate Management Admission Test), regeln den Zugang von Graduierten. In zahlreichen Validitätsstudien konnte die prognostische Validität dieser Verfahren aufgezeigt werden, obwohl sie trotz des immensen Aufwands einen großen Teil der auftretenden Varianz nicht erklären können (Schmitt et al., 2007). Burton und Wang (2005) berichten eine unkorrigierte Korrelation zwischen dem GRE und den Studiennoten im Graduiertenstudium von $r = 0,40$ beziehungsweise korrigiert¹¹ von $r = 0,49$. Wurden als Kriterien Ratings der Dozenten verwendet, so zeigten sich leicht höhere Zusammenhänge. Kuncel, Hezlett und Ones (2001) führten eine unabhängige Metaanalyse durch und kamen zu vergleichbaren Ergebnissen: Die prädiktive Validität der gleichgewichteten Faktoren verbal, quantitativ und analytisch wird mit $r = 0,45$ berechnet. Die hohe Korrelation zwischen diesen Prädiktoren und den College-Studiennoten zeigt sich eindrücklich, wenn deren inkrementelle Validität betrachtet wird. Die Aufnahme der College-Studiennoten erreicht lediglich eine Verbesserung von $\Delta r = 0,04$. Kritischer betrachten Sternberg und Williams (1997) den GRE. Sie beanstanden gerade die theorielose Entwicklung des Verfahrens und weisen auf die teilweise geringen Validitäten bei einzelnen Personengruppen hin.

Die Validität des SAT I – also des fachunabhängigen Reasoning-Tests des Educational Testing Service für den College-Eintritt – wurde ebenfalls in zahlreichen Studien überprüft (zum Beispiel Bridgeman, McCamley-Jenkins & Ervin, 2000; Burton & Ramist, 2001). Auf der Basis von 16 995 beziehungsweise von 82 000 Studenten in den USA wurde ein unkorrigierter Validitätskoeffizient zur Vorhersage von Studiennoten $r = 0,36$ beziehungsweise zur Vorhersage der Persistenz $r = 0,33$ berechnet. Eine Metaanalyse von nahezu 3 000 Studien zeigte ebenfalls, dass der SAT ein valider Prädiktor für die frühen Studienleistungen ($r = 0,44$ bis $0,62$) ist (Hezlett et al. 2001, zitiert nach Kobrin, Camara & Milewski, 2002). Zudem konnte nachgewiesen werden, dass auch eine Vorhersage der späteren Studienleistungen (kumulative Studiennoten, Persistenz versus Studienabbruch) mit Validitätskoeffizienten von etwa $r = 0,35$ bis $0,45$ möglich ist. Allerdings wird in den letzten Jahren verstärkt über die Fairness des SAT diskutiert: So konnte ein Effekt des sozioökonomischen beziehungsweise ethnischen Hintergrunds nachgewiesen werden (Everson, 2004; W. E. Sedlacek, 2004).

Der Test für medizinische Studiengänge, entwickelt durch das Institut für Test- und Begabungsforschung, heute durch die ITB Consulting GmbH, regelte über 16 Jahre die Studienplatzvergabe in der Medizin an deutschen Hochschulen, später auch als „Eignungstest für das Medizinstudium“ (EMS) in der Schweiz und in Österreich. So liegt die Korrelation des TMS-Testergebnisses mit dem Gesamtwert der Ärztlichen Vorprüfung bei $r = 0,48$, die des Abiturs bei $r = -0,45$ (Stumpf & Nauels, 1990). Die Korrelationen mit der ersten ärztlichen Prüfung fallen mit $r = 0,33$ beziehungsweise mit $r = -0,32$ geringer aus. Bei der gewichteten Kombination von Test und Abitur liegt der durchschnittliche Prognosekennwert bei $r = 0,51$. Vergleichbare

¹¹ Die Ergebnisse sind nur korrigiert für Varianzeinschränkung.

Ergebnisse wurden in der Schweiz erzielt (Hänsgen & Spicher, 2002). Studierende, die über dieses Kombinationskriterium zugelassen wurden, traten im Mittel nach kürzerer Studiendauer zur Ärztlichen Vorprüfung an. Gleichzeitig waren sie beim ersten Anlauf in höherem Maße erfolgreich als Studierende, die über andere Zugangswege einen Studienplatz erhalten hatten (Trost, Klieme & Nauels, 1997). Eindrücklich zeigt sich eine Erfolgsquote für die Testbesten im TMS von 97 %. Für Personen, die aufgrund des leistungsgesteuerten Losverfahrens zugelassen wurden, wurde eine Quote von 71 % kalkuliert (Deidesheimer Kreis, 1997). Insgesamt wurde mit dem TMS eine höhere Trefferquote erreicht als mit dem Allgemeinen Auswahlverfahren.

In zahlreichen Studien konnte also nachgewiesen werden, dass die gerade vorgestellten Testverfahren die Prognosefähigkeit des Abiturs beziehungsweise des High-School-GPA verbessern. Ob der Einsatz solch differenzierter Verfahren – wie eben der SAT oder GRE – jedoch die Prognosefähigkeit im Vergleich zu allgemeinen Intelligenzverfahren verbessert, ist Gegenstand heftiger wissenschaftlicher Dispute: Die Arbeitsgruppe von Ree konnten im Bereich Arbeitsleistungen eine Verbesserung der Prädiktion korrigiert¹² im Mittel von $\Delta r = 0,06$ (Ree, Earles & Teachout, 1994) und im Bereich Ausbildungserfolg von 0,02 (Ree & Earles, 1991) nachweisen. Olea und Ree (1994) kalkulierten bei den Ausbildungsleistungen von Piloten ein Inkrement von $\Delta = 0,08$ und von Flug navigatoren von 0,02. Schmidt-Atzert, Deter und Jaeckel (2004) verwandten für eine ähnliche Fragestellung im deutschen Ausbildungssystem den Wilde-Intelligenz-Test von Jäger und Althoff (1983). Bei vier technischen und zwei kaufmännischen Berufen zeigte sich dabei die Allgemeine Intelligenz nahezu durchgehend als bester Prädiktor. Von den fünf primären kognitiven Fähigkeiten nach Thurstone leistete für die Vorhersage theoretischer Kenntnisse nur die Dimension des räumlichen Vorstellungsvermögens in einem von insgesamt sechs Berufen einen zusätzlichen signifikanten Beitrag. Bei der Vorhersage von praktischen Fertigkeiten konnten immerhin von sechs Berufen drei gefunden werden, in denen die Vorhersage signifikant verbessert werden konnte. In einer aktuellen Metaanalyse konnten Brown, Le und Schmidt (2006) dieses Ergebnis bestätigen: Die Neuentwicklung spezifischer Fähigkeitstests, aber auch eine differenzierte Gewichtung bestimmter Intelligenzdimensionen bringen diesen Autoren zufolge vergleichsweise geringe bis keine inkrementelle Validität – insbesondere auch dann, wenn Kriterium und Prädiktor minderungskorrigiert werden.

Ackerman und Cianciolo (2002) führen zu dieser generellen Aussage der Arbeitsgruppen von Ree, Schmidt und Hunter an, dass bei den meisten entsprechenden Studien Testverfahren (meist die Armed Forces Vocational Aptitude Battery beziehungsweise die General Ability Test Battery) verwandt wurden, die nicht dazu geeignet sind, umfassend differentielle Intelligenzkomponenten abzubilden. Zudem scheint der Abstraktionsgrad des Kriteriums sowie des Prädiktors für die Prädiktionsleistungen entscheidend zu sein (Wittmann & Matt, 1986; Wittmann & Süß, 1999), so dass die Ergebnisse nicht verallgemeinert werden können: Dabei hat jede Ebene der Generalität auf der Seite des Prädiktormodells eine entsprechend symmetrische

¹² Die Ergebnisse sind nur korrigiert für Varianzeinschränkung.

Ebene der Generalität auf Kriterienebene und umgekehrt. Übertragen auf die Studienauswahl sind generelle Kriterien wie zum Beispiel die Studienzufriedenheit oder aber auch die allgemeine Arbeitsleistung durch eher umfassendere, generellere Prädiktoren vorhersagbar. Spezifische Kriterien, zum Beispiel das Abschneiden in einer bestimmten Klausur in Mathematik, erfordern auch spezifische Prädiktoren. Verallgemeinert bedeutet dies, dass die generelle mentale Fähigkeit zwar häufig, aber – abhängig vom Kriterium – nicht *immer* den besten Prädiktor darstellt.

Ein weiterer wesentlicher Aspekt beschäftigt sich damit, inwieweit überhaupt Studierfähigkeitstests spezifischere Anforderungen abbilden als allgemeine Intelligenztests. Auch wenn die Entwicklung eines (vielleicht gar fachspezifischen) Studierfähigkeitstests zu Itemmaterial führt, das passend zum Studienfach kontextualisiert ist, bedeutet dies noch lange nicht, dass dieser nicht nur fast ausschließlich g misst. Frey und Detterman sind dieser Frage nachgegangen und berichten von Korrelationen des SAT, einem allgemeinen Studierfähigkeitstest, mit der Generellen Mentalen Fähigkeit von $r = 0,86$ (gemessen mit der Armed Services Vocational Aptitude Battery, Nichtlinearität korrigiert) beziehungsweise von $r = 0,72$ (Raven's Advanced Progressive Matrices, Varianzeinschränkung korrigiert) (Frey & Detterman, 2004a). Demgemäß veröffentlichten sie eine Regressionsfunktion, die Testergebnisse des SAT in Intelligenzquotienten konvertiert (Frey & Detterman, 2004b). Coyle (2006) wies analog dazu nach, dass sowohl die Ergebnisse von SAT als auch von ACT-Testungen hoch auf g laden. Empirisch scheinen also zumindest die beiden international einflussreichsten Studierfähigkeitstests eine hohe Ähnlichkeit mit allgemeinen Intelligenztests aufzuweisen. Vergleichbare Ergebnisse lassen sich aber auch zum Medizinertest aufzeigen: Trost (2003) berichtet eine unkorrigierte Korrelation von $r = 0,70$ des Tests für medizinische Studiengänge, einem fachspezifischen Studierfähigkeitstest, mit einem allgemeinen Intelligenztest.

Diese letzten Ergebnisse lassen den Schluss zu, dass Studierfähigkeitsverfahren mit Tests zur allgemeinen Intelligenz so hoch korrelieren, dass sich eine spezielle Entwicklung nicht lohnen würde. Unbeachtet blieb jedoch bis hierhin, dass im Bereich der Studieneignung und ganz besonders der Studienberatung differentielle Aussagen vonnöten sind, soll doch nicht nur eine reine Selektion, sondern auch eine Platzierung auf Basis der differenzierten Betrachtung unterschiedlicher Person-Studienfach-Passung durchgeführt werden (Cronbach & Gleser, 1965). Solche differenzierten Prognosen können ausschließlich mit einer spezifischen Testbatterie getroffen werden, die die einzelnen Intelligenzkomponenten getrennt betrachtet: So zeigte sich in den letzten Jahren zum Beispiel im Bereich der Frühförderung von Hochbegabten eine Verschiebung der Hauptforschungsrichtung von „an emphasis primarily on general cognitive ability to an appreciation of the unique information afforded by verbal and quantitative abilities“ (Shea, Lubinski & Benbow, 2001, S. 604). Mit einer getrennten Erfassung verbaler, räumlicher und numerischer Fähigkeiten konnte entsprechend nachgewiesen werden, dass Personen mit besonderer Begabung in der Mathematik sich anders entwickelten als Begabte im sprachlichen Bereich –

trotz einer gleich hohen allgemeinen Intelligenz (Lubinski, 2004). Shea, Lubinski und Benbow (2001) identifizierten bei Jugendlichen von 13 Jahren mit einer Testbatterie zur *GMA* 1 % der Testbesten. Gleichzeitig wurden Tests zum quantitativen, räumlichen und verbalen Reasoning durchgeführt. Gerade die spezifischen Fähigkeiten schlugen sich in der Präferenz für bestimmte Schul- beziehungsweise Studienfächer und Berufe nieder.

Damit ist aber noch nicht belegt, dass die spezifische Studierfähigkeit einen grundsätzlich anderen Intelligenztest als die Messung der spezifischen Intelligenz erfordert (vgl. auch Kersting, 2005). Die oben bereits genannte Studie von Rosas Díaz (1990) konnte nämlich mit dem Berliner Intelligenz-Strukturmodell bei Berücksichtigung der einzelnen BIS-Komponenten eine differentielle Eignung für verschiedene Studiengänge prognostizieren. Einzelne gewichtete Komponenten zeigen eine höhere Korrelation mit dem Studienerfolg als die allgemeine Intelligenz: So konnte zum Beispiel Kersting (1999) in einer Studie zur Zulassung zum Finanzwirt-Studium die besondere Bedeutung der Verarbeitungskapazität numerischer Inhalte nachweisen.

Der größte Unterschied zwischen einem speziell gewichteten Intelligenztest und einem speziellen Studierfähigkeitstest sind nicht die abgebildeten Konstrukte, sondern vielmehr die besondere Kontextualisierung. Genau diese Einbettung der zu messenden Sachverhalte in das textliche Umfeld verbessert nicht unbedingt die Validität, verteuert die Entwicklung, aber steigert die Akzeptanz des Verfahrens bei Studieninteressierten und Hochschuldozenten (Hell & Schuler, 2005). Schülern oder auch Dozenten gelingt es anscheinend besser zu erraten, was die jeweiligen Aufgaben mit dem Studienerfolg zu tun haben, und sie vertrauen dem Verfahren so eher als anderen, generelleren Verfahren.

3.4 Nicht-kognitive Verfahren als Prädiktoren

Auch wenn sich – wie im vorigen Kapitel resümiert – in empirischen Studien die Bedeutsamkeit gerade der kognitiven Prädiktoren (allen voran der allgemeinen Intelligenz) zeigt, zeichnet sich bei Anforderungsanalysen immer wieder die besondere Bedeutung nicht-kognitiver Prädiktoren für die Prädiktion des Studienerfolgs ab (Heldmann, 1984; Konegen-Grenier, 2001; Zimmerhofer, 2003). Es ist nachvollziehbar, dass nicht nur die allgemeine Fähigkeit, Zusammenhänge schnell zu begreifen und entsprechend neue Problemlagen auf Basis bereits Gelerntem zu lösen, erfolgskritisch ist. Gerade aufgrund der Freiheit an den Universitäten mit dem für Studienanfänger niedrigen Grad der Verschulung sind Aspekte der Motivation für einen erfolgreichen Studienabschluss augenscheinlich. Zudem stellen Studierende bereits eine nach der Intelligenz ausgewählte Subpopulation dar, so dass gerade Persönlichkeit und Motivation gut geeignet sein könnten, Unterschiede zwischen der scheinbar homogenen Gruppe aufzudecken (Cattell, 1965; Trost, 1975).

In den USA wurden lange Jahre nahezu ausschließlich standardisierte Studierfähigkeitstests – allen voran der SAT des Educational Testing Service, Princeton und der ACT des gleichnamigen Instituts (ehemals American College Testing) in Iowa City – eingesetzt. In den letzten Jahren stieg allerdings das Interesse an alternativen Prädiktoren des Studienerfolgs. Allen voran der Präsident der Universität Kalifornien, Richard Atkinson (2002), und William Sedlacek (2001, 2004) brachten eine Debatte um diese standardisierten Studierfähigkeitstests in die Öffentlichkeit. Hauptkritikpunkte an den oben genannten Verfahren beziehen sich auf der einen Seite auf die geringe Fairness bezüglich Minderheiten und auf der anderen Seite auf die gute Trainierbarkeit durch kostenpflichtige Vorbereitungskurse. Auch wird die geringe inkrementelle Validität zum Beispiel des SAT I über die High-School-Abschlussnoten angemerkt. Als Antwort auf diese Kritik steigt die Bereitschaft, die Rolle von Faktoren abseits der allgemeinen oder studienfeldspezifischen Studierfähigkeit zu evaluieren und besonders den Einsatz nicht-kognitiver Faktoren zu bedenken.

3.4.1 Big-Five-Persönlichkeitsdimensionen

In den letzten Jahren ist das Fünf-Faktoren-Modell oder kurz Big-Five (Goldberg, 1990; McNulty, 2000) zum einflussreichsten Faktorenmodell der Persönlichkeit avanciert. Heute werden entsprechende Fragebogen in vielen Forschungs- aber auch Anwendungsgebieten, zum Beispiel in der Wirtschaft, verwandt. Damit ist eine unüberschaubare Vielzahl von einzelnen, unterschiedlich benannten und konzipierten Persönlichkeitsfaktoren durch ein effizientes und international anerkanntes Faktorenmodell abgelöst worden.

Umschreiben lassen sich die fünf Faktoren wie folgt (Rammstedt, Koch, Borg & Reitz, 2004, S. 28):

1. *Extraversion*: gesprächig, gesellig, voller Tatendrang, kontaktfreudig versus still, schweigsam, zurückgezogen, eher ein Einzelgänger
2. *Verträglichkeit*: kooperativ, herzlich, nett, nachgiebig versus kühl, kritisch, wird leicht ärgerlich, misstrauisch
3. *Gewissenhaftigkeit*: zuverlässig, sorgfältig, ordentlich, pflichtbewusst, ehrgeizig versus salopp, unordentlich, unpünktlich, chaotisch, nonchalant
4. *Neurotizismus oder emotionale Stabilität*: nervös, ängstlich, regt sich leicht auf, angespannt, empfindlich versus gelassen, entspannt, selbstzufrieden, robust
5. *Offenheit für Erfahrung oder auch Intellekt*: offen für Neues, wissbegierig, kultiviert, phantasievoll versus feste Ansichten und Meinungen, an Neuem eher weniger interessiert, traditionell, mag Routinen und feste Regeln

Seit den 1990er Jahren sind international zahlreiche Studien veröffentlicht worden, die einen statistischen Zusammenhang zwischen der Big-Five-Persönlichkeitsstruktur und beruflicher

Leistung untersuchen. So konnten Barrick und Mount (1991) beziehungsweise Tett, Jackson und Rothstein (1991) in Metaanalysen die Nützlichkeit von Persönlichkeitsfragebogen auch für die Personalauswahl belegen. Besonders die Big-Five-Dimensionen zeigten sich über europäische (Salgado, 1997; Salgado, 1998) und US-amerikanische Studien (Barrick, Mount & Judge, 2001; Hurtz & Donovan, 2000) hinweg als gute Prädiktoren für die Leistung in verschiedenen Berufen, verschiedenen Organisationen und in verschiedenen Ländern dieser Erde. Auch wenn die entsprechenden korrigierten Korrelationen mit meist $r < 0,3$ lediglich im mittleren Bereich für statistische Zusammenhänge liegen, zeigen insbesondere die emotionale Stabilität und die Gewissenhaftigkeit eine inkrementelle Validität über die allgemeine Intelligenz hinaus (Salgado, 1998). Der Forschung zufolge erreichen im direkten Vergleich von Arbeitnehmern mit denselben kognitiven Voraussetzungen diejenigen höheres Berufswissen, die ihre Arbeit gewissenhafter verrichten. Diese Personen investieren größere Anstrengungen und mehr Zeit für die Bearbeitung von Aufgaben – mit entsprechenden Auswirkungen auf die Arbeitsleistung.

Dabei steigert die Verwendung der Persönlichkeitsfaktoren über Berufe und Kriterien hinweg die Validität um etwa 10 % (in einem linearen Modell), mit einem Schwankungsbereich je nach Berufstätigkeit und Kriterium zwischen 7 % und 38 % (Schmidt & Hunter, 1998). Entsprechend gehören Persönlichkeitsfragebogen zu den geeignetsten Prädiktoren in der Berufseignungsdiagnostik und sind damit valider als zum Beispiel Interviews oder Assessment-Center (Schmidt & Hunter, 1998). Mithilfe von Persönlichkeitsfragebogen lässt sich besonders gut – gar besser als mit der Allgemeinen Intelligenz (Judge, Higgins, Thoresen & Barrick, 1999) – eine Vorhersage der späteren Arbeitszufriedenheit im Längsschnittdesign erreichen ($R = 0,42$, unkorrigiert). Metaanalytisch konnten Judge, Heller und Mount (2002) Korrelationen der subjektive Zufriedenheit mit der Dimension emotionale Stabilität ($r = -0,29$, korrigiert¹³), mit Extraversion ($r = 0,25$), mit Verträglichkeit ($r = 0,17$) und mit Gewissenhaftigkeit ($r = 0,26$) nachweisen.

Die hier in aller Kürze berichteten Studien deuten insgesamt darauf hin, dass es sich durchaus lohnt, neben Intelligenz- auch Persönlichkeitsdimensionen in der Personaldiagnostik einzusetzen, wäre das Problem der absichtlichen Verfälschbarkeit bei letzteren gelöst (Diemand & Schuler, 1991, Zimmerhofer & Hornke, 2005a). Fraglich ist aber an dieser Stelle, inwieweit diese Ergebnisse auf den Bereich Schule und Studium übertragen werden können. Nicht zu leugnen ist dabei, dass ein durchaus großer Unterschied besteht zwischen dem Berufs- und dem Studienalltag und besonders hinsichtlich deren Erfolgskriterien. Aber auch für Lernsituationen konnte der Zusammenhang mit Persönlichkeitsdimensionen belegt werden. Dabei scheinen insbesondere Extraversion, Gewissenhaftigkeit und Offenheit für Erfahrungen relevant zu sein (De Raad & Schouwenburg, 1996); entsprechend werden diese Dimensionen *educational circumplex* (De Raad, 1996, S. 197) genannt. Dabei ist – genauso wie im Berufsalltag – die Gewissenhaftigkeit allein der beste Prädiktor für akademischen Erfolg (Busato, Prins, Elshout & Hamaker, 2000).

¹³ Die Ergebnisse sind doppelt minderungskorrigiert mit Berücksichtigung der Varianzeinschränkung.

Pfadanalysen zum Zusammenspiel von Persönlichkeit und Studien- beziehungsweise Schulleistung zeigen, dass Lernstrategien durch Persönlichkeitsvariablen vorhergesagt und dass akademischer Erfolg wiederum durch Lernstrategien beeinflusst werden (Diseth, 2003). Somit ist das Verhältnis von Persönlichkeit und Leistungsnoten durch die verwandten Lernstrategien mediiert (Blickle, 1996). Lernstrategien, die auf ein tieferes Verständnis des Stoffes abzielen (*deep approach*), korrelieren positiv mit Offenheit für Erfahrungen, ein oberflächliches Lernen mit Neurotizismus, während eine strategische Prüfungsvorbereitung (Kombination aus tiefem und oberflächlichem Lernen – je nach Situation) mit Gewissenhaftigkeit (Diseth, 2003) zusammenhängt. Auch wenn der *deep approach* beim Lernen allgemein von Vorteil ist (Diseth, 2003; Newstead, 1992; Sadler-Smith, 1997), kann er sich im Einzelfall dysfunktional auf Prüfungen auswirken (Wild, Krapp & Winteler, 1992), die einen hohen Reproduktionsanteil besitzen. Oberflächliche Strategien sind dagegen klarer mit schwächeren Prüfungsergebnissen verknüpft. Auch unstrukturiertes Vorgehen (*undirected learning style*) zeigt sich als konsistenter negativer Prädiktor für akademischen Erfolg (Busato, Prins, Elshout & Hamaker, 2000). Diese Form der Prüfungsvorbereitung korreliert zudem negativ mit Gewissenhaftigkeit ($r = -0,09$) und Offenheit für Erfahrungen ($r = -0,17$), hingegen positiv zu $r = 0,21$ mit Neurotizismus (Busato, Prins, Elshout & Hamaker, 1999). Insgesamt werden Varianzaufklärungen (Studiennoten als Kriterium) durch Lernstrategien und Persönlichkeitsdimensionen von bis zu 30 % aufgezeigt (Blickle, 1996).

In einer Metaanalyse zur Vorhersage des Studienerfolgs zeigten sich erwartungsgemäß Gewissenhaftigkeit (korrigierte¹⁴Korrelation $r = 0,25$) und Offenheit für Erfahrungen (korrigierte Korrelation $r = 0,13$) als wichtigste Prädiktoren (Trapmann, Hell, Hirn, Weigand & Schuler, 2005). Dabei wurden als Kriterium – wie im Allgemeinen üblich – die Studiennoten gewählt. Die Persistenz lässt sich durch die emotionale Stabilität nur eingeschränkt vorhersagen (korrigierte¹⁴Korrelation $r = 0,06$). Eine höhere Varianzaufklärung wird hingegen in dieser Studie für das Kriterium Studienzufriedenheit erreicht: Die emotionale Stabilität erreicht dort $r = 0,37$, einzelne Facetten der Gewissenhaftigkeit (Leistungsstreben und Selbstdisziplin) vergleichbare Ergebnisse. Allerdings stehen in diesem Bereich nur wenige Primärstudien – anders als zur Prädiktion der Studiennoten – zur Verfügung.

3.4.2 Rubikon-Modell der Handlungsphasen

Im Folgenden wird das Rubikonmodell der Handlungsphasen (Gollwitzer, 1996) als ein deskriptives Modell vorgestellt, dass sowohl die Motivationsprozesse als auch so genannte Volitionsprozesse näher betrachtet. Verwandt wird dieses Modell hier, um die einzelnen Konzepte von beruflichem Interesse (siehe Abschnitt 3.4.3), von extrinsischer Lernmotivation (3.4.4), von

¹⁴ Die Ergebnisse sind korrigiert für den Messfehler im Kriterium, nicht im Prädiktor.

Selbstwirksamkeit (3.4.5) und von Handlungs- versus Lageorientierung (3.4.6) integrieren zu können.

Der Name des Modells befasst sich direkt mit dem Zusammenspiel, genauer gesagt, mit dem Übergang zwischen Motivation und Volition. Den Namen Rubikon bezieht Gollwitzers und Heckhausens Modell aus einer geschichtlichen Episode aus dem Jahr 49 vor Christus. Caesar überschritt in diesem Jahr nach langem Abwägen den „Rubikon“, der südlich von Ravenna in die Adria mündet und soll dabei die bekannten Worte „Alea iacta est“ gesagt haben. Der Würfel war gefallen; es gab also kein Zurück mehr: Aus dem Wunsch wurde eine klare Intention. Das Modell beschreibt einen Prozess aus vier Phasen, die von der Planung zur Ausführung einer Handlung führen.

Phase 1: In Heckhausens und Gollwitzers Modell (1996) beginnt der Prozessablauf mit einer Motivationsphase (siehe Abbildung 3.1 zur Verdeutlichung). Dabei werden einzelne Wünsche wahrgenommen, die bei dem Individuum bestehen, in ihrem Wert und hinsichtlich ihrer Machbarkeit beurteilt. Entscheidend dabei ist das eigene Interesse (siehe Abschnitt 3.4.3). Auch wird geklärt, ob sich das Ergebnis ohne Zutun einstellt und wie wahrscheinlich das Ergebnis je nach eigener Handlungen ist (Rheinberg, 1995). Die Wahrnehmung eigener Selbstwirksamkeit (siehe Abschnitt 3.4.5) ist entscheidend. Ferner werden aber auch die Folgen des Ergebnisses und deren Bedeutsamkeit betrachtet. Diese Phase beschäftigt sich mit der Aufbereitung der Informationen für die weitere Entscheidung: Für Studenten könnte das bedeuten, dass sie sich überlegen, was sie am Vormittag machen. Gehen sie zur Vorlesung? Wie wichtig ist diese wirklich für die Prüfung? Lernen sie lieber zu Hause oder entspannen sie sich lieber? Hier kann es zu widerstrebenden Anreizen kommen: Förderlich für die Lernhandlung sind die Antizipationen der Folgen – und zwar getrennt nach der eigenen Bewertung und der Fremdbewertung. Folge einer guten Prüfungsvorbereitung und guter Prüfungsleistungen sind also sowohl der Stolz über die eigene Leistung als auch die Anerkennung anderer, die einem widerfährt. Weiter förderlich ist aber natürlich auch das Interesse, den Stoff näher kennen zu lernen. Neben den Einflüssen, die eine Entscheidung für die Vorlesung begünstigen, sind auch verschiedene Faktoren denkbar, die gegen diese Entscheidung sprechen. Vertrauen die Personen ihrer Leistung nicht, bewerten sie jedwede Lernanstrengung als aussichtslos. Hier greift wiederum die Selbstwirksamkeit in den Prozess ein (siehe Abschnitt 3.4.5). Vielleicht zeigt die Person aber auch Prüfungsangst und möchte daher am liebsten überhaupt nicht an die Prüfung erinnert werden. Daher setzt sie sich möglicherweise in Folge auch dem Lernen nicht aus. Der Abschluss der ersten Phase wird durch einen Entschluss erreicht. In diesem hier skizzierten Fall möge entschieden worden sein, zu den Vorlesungen zu gehen und sich verstärkt auf die Prüfung vorzubereiten. Damit ist der Rubikon überschritten.

Phase 2: Um auch im Modell abbilden zu können, dass Intentionen (also eben der Plan, sich verstärkt vorzubereiten) nicht immer direkt realisiert werden, wird von den Autoren eine

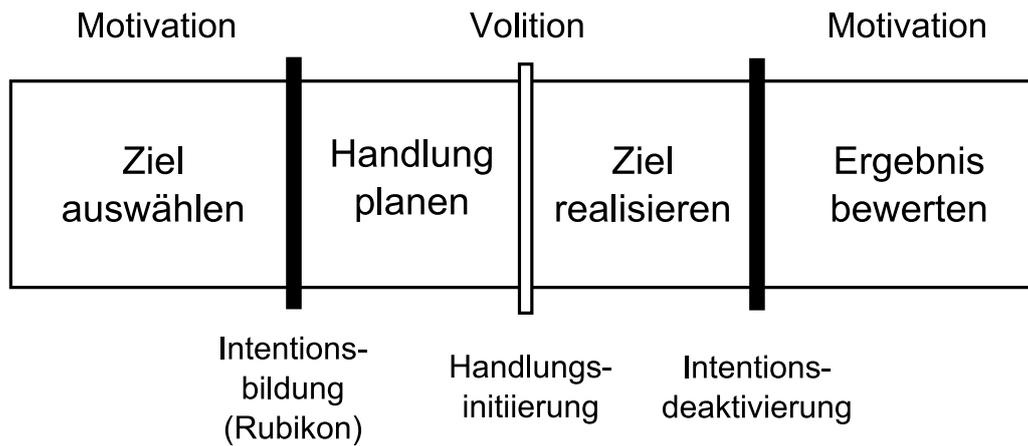


Abbildung 3.1: Das Rubikonmodell des Handelns (modifiziert nach Achtzinger und Gollwitzer, 2006)

Unterteilung in die präaktionale und aktionale Volitionsphase vorgenommen. In der präaktionalen Phase steht die Konkretisierung der Intention im Mittelpunkt. Dabei wird die Planung des *Wann*, *Wie* und *Wo* lange ausgearbeitet. Gerade im Fall einer Procrastination, also des Aufschiebens von Prüfungsleistungen, kann es auch dazu kommen, dass die Personen durch exzessive Planungen vergleichsweise kleinerer Aufgaben, unnötig detaillierte Zeitpläne und extrem ausführliche Gliederungen die spätere Durchführung eher erschweren als unterstützen. Hier wird ein Bereich der Handlungsorientierung bei Tätigkeitsplanung näher betrachtet werden müssen (siehe Abschnitt 3.4.6).

Phase 3: Zwischen den beiden Phasen der Volition liegt der Entschluss, die Handlung in dieser Situation zu initiieren. Die so genannte Fiat-Tendenz entscheidet dabei über den Start einer Handlung, die wiederum abhängig ist von der Einschätzung, welche Intention in der jetzigen Situation besonders passt und was als besonders dringlich erachtet wird. Die Handlungsphase selbst besteht aus der aktionalen Volitionsphase. Dabei können bereits begonnene Handlungen gestört, abgebrochen oder unterbrochen werden. Selbstzweifel, zu geringes Durchhaltevermögen, aber auch der Einfluss anderer, attraktiverer Alternativen beeinflussen den weiteren Verlauf. Wichtig dabei sind Selbstkontrolltechniken, die einer Person ermöglichen, auch bei attraktiven Alternativen die Handlungsausführung zu kontrollieren und somit das gesetzte Ziel nicht aus den Augen zu verlieren. Die Handlungsorientierung bei Tätigkeitsausführung und nach Misserfolg wird hier näher betrachtet werden müssen (siehe Abschnitt 3.4.6).

Phase 4: Nach der Phase 3 beginnt gemäß dem Modell wieder eine Motivationsphase, die postaktional genannt wird. Dabei wird eine entsprechende Begutachtung des erreichten Ziels vorgenommen. Entsprechend wird dann – wenn das Ziel verfehlt wurde – neu entschieden, ob und wie das Ziel weiter verfolgt werden soll.

In Strukturgleichungsmodellen zeigen sich zwei unabhängige kausale Pfade zur Aufklärung von Studienleistungen (Schiefele & Urhahne, 2000): Die motivationale Komponente, zu der in diesem Kapitel sowohl das Interesse (siehe Abschnitt 3.4.3) als auch die extrinsische Lernmotiva-

tion (siehe Abschnitt 3.4.4) gehören, und die Volition (siehe Abschnitt 3.4.5 und 3.4.6) konnten dabei unterschieden werden. Zudem scheint der Notendurchschnitt stärker von motivationalen (Studieninteresse und Leistungsmotivation) als von volitionalen Faktoren (Handlungskontrolle und Selbstwirksamkeit) abzuhängen, die Menge der Klausuren oder die Studiendauer eher von der Volition.

3.4.3 Berufliche Interessen

Gerade für die Studienfachwahl spielt das Interesse eine zentrale Rolle, messen doch die meisten Studenten diesem Einflussfaktor zumindest subjektiv die entscheidende Bedeutung bei ihrer Studienfachwahl zu (Dawis & Lofquist, 1984; Lewin, Heublein, Schreiber, Spangenberg & Sommer, 2002; Heublein & Sommer, 2002). Diesem zum Trotz, geben eine Mehrzahl von Befragten zu, dass sie vergleichsweise wenig über die einzelnen Studienfächer wissen (Heublein & Sommer, 2002) – meist zu wenig, um auf Basis ihrer Interessen eine zuverlässige Wahl treffen zu können. Entsprechend scheinen Personen, die die Studienberatung konsultieren, weniger zu ihren Fähigkeiten, sondern verstärkt zu den eigenen Interessen und zu den entsprechenden Befriedigungspotenzialen möglicher Studienfächer beraten werden zu wollen (Bergmann, Brandstätter & Eder, 1994; Brandstätter, Farthofer & Grillich, 2001).

Um mehr Klarheit im Umgang mit dem vorwissenschaftlichen und dem wissenschaftlichen Begriff *Interesse* zu schaffen, werden im Folgenden zwei Herangehensweisen unterschieden: Eine direkt erfragte Fachpräferenz gleicht den so genannten *expressed interests* (Silvia, 2001). Dabei werden direkte, offene Fragen bezüglich des Studienwunsches gestellt. Im einfachsten Fall wird eine Formulierung benutzt, wie z.B. „Welche Fachrichtung wollen Sie studieren?“ (siehe den benutzten Fragebogen in Heine, Scheller & Willich, 2005). Der Einzelfall für die Beratung steht so klar im Vordergrund und somit der ideographische Ansatz (Silvia, 2001). Wie beschrieben, ist es für Heranwachsende jedoch nicht leicht, einzelne Studienfächer aus dem riesigen Angebot zu präferieren und präzise eine Studienfachwahl anzugeben. Dies ist aber bei dem obigen Ansatz notwendig. Helfen können die *measured interests*. Dabei messen auf Interessentheorien basierte, standardisierte Fragebogen mithilfe zu bewertender Tätigkeiten die Interessenstruktur und Passung zur Umwelt. Unentschlossene können somit geleitet werden, ohne eine direkte Präferenz für Studienfächer angeben zu müssen. Bei einem entsprechenden Fragebogen (daher ist auch das Synonym *inventoried interests* geläufig) werden die Präferenzaussagen bezüglich der Tätigkeiten ausgewertet und meist mit einer Normstichprobe verglichen: Dies entspricht dem nomothetischen Ansatz. Berechnungen des Zusammenhangs zwischen Personen-Umwelt-Kongruenz und Studienerfolg (Bergmann, 1992) oder der späteren Berufswahl (Tracey & Hopkins, 2001) sind genauso durchführbar wie die Überprüfung der Faktorenstruktur des Fragebogens.

In zahlreichen Studien konnte der geringe Zusammenhang zwischen der Befragung mit *expressed* und *measured interests* des Fachinteresses nachgewiesen werden (Bartling & Hood, 1981; Borgen & Seling, 1978). Es zeigt sich, dass das artikulierte Interesse bezüglich der Einmündungsvorhersage¹⁵ zunächst valider ist. Für den weiteren Studienverlauf (also Stabilität und Zufriedenheit) ist aber das gemessene Interesse der wirksamere Prädiktor als die artikulierte Studienfachwahl (Bergmann, 1994). Ferner ist gerade eine Diskrepanz zwischen den gemessenen und den artikulierten Interessen ein deutliches Warnsignal für weiteren Beratungsbedarf. Diese Gründe sprechen also für eine weitere Betrachtung der anhand von einer Interessentheorie geleiteten Analyse der individuellen Studienpräferenz.

Auch wenn der Begriff *Interesse* hier bereits häufig benutzt wurde und in der Alltagssprache allgegenwärtig ist, scheint eine wissenschaftliche Betrachtung unumgänglich, um das umgangssprachliche Verständnis des Wortes zu korrigieren:

Das Wort *Interesse* stammt vom lateinischen *inter* (zwischen) und *esse* (sein) ab. Schon diese Übersetzungen deuten auf *Interesse* als einem relationalen Verhältnis zwischen zwei Elementen hin. Jacob und Wilhelm Grimm beschrieben bereits im 19. Jahrhundert *Interesse* als „wohlgefallen [...], das wir mit der vorstellung [sic] der existenz [sic] eines gegenstandes [sic] verbinden.“ (Grimm & Grimm, 1854-1960). Diese Relation zwischen einer Person und einem Gegenstand ist bis heute Bestandteil nahezu aller Interessentheorien. Der Gegenstand wird dabei – so zum Beispiel in der pädagogischen Interessentheorie sensu Prenzel, Krapp und H. Schiefele (1986) – „als Umweltausschnitt definiert, den die Person von anderen Umweltbereichen unterscheidet und als eingegrenzte und strukturierte Einheit abbildet“ (S. 166). Unterschieden werden zwei Formen dieses *Interesses*: Das so genannte *situationale oder auch aktualisierte Interesse* wird nicht als eine Persönlichkeitsdisposition verstanden, da die Zugewandtheit zu einem bestimmten Gegenstand nicht eine längere Zeit überdauert. Essenzieller für eine Betrachtung hier ist das *individuelle oder dispositionale Interesse*, das im Gegensatz dazu persönlichkeitspezifische Wertvorstellungen und Handlungsbereitschaften widerspiegelt. Diese sind im Wertesystem des Individuums verankert und bilden entsprechend einen Teil des Selbstkonzepts. Damit bestimmen sie die Entscheidung des Individuums über die Tagesgestaltung, also auch über das Lern- und Prüfungsverhalten.

Der Zusammenhang von *Interesse* und Lernmotivation scheint somit klar: Die motivationale Dynamik eines *Interesses* rührt daher, dass sich eine Person mit den Inhalten und Tätigkeiten eines Interessenbereichs – zum Beispiel eines Studienfachs – möglichst überdauernd identifiziert, dieses also für persönlich wichtig erachtet. Wählen Schüler ihren Interessen zufolge Studienfächer beziehungsweise Vorlesungen oder Seminare, so befassen sie sich gerne mit den geforderten Lernaufgaben, werden entsprechend mehr Anstrengung zeigen, generieren Wissensbestände und entwickeln Kompetenzen (Schiefele, Wild & Winteler, 1995; Winteler, Sierwald & Schiefele,

¹⁵ Dies ist die Vorhersage der Studienfachwahl der Person. Der weitere Studienverlauf, also zum Beispiel die Persistenz, wird dabei nicht betrachtet.

1988; Wild, Krapp & Winteler, 1992). Zentral dabei ist aber, dass die Handlungsveranlassung als freiwillig, dass heißt als selbst gewollt, erlebt ist (Deci & Ryan, 1993). Aus dem Interesse und der selbst gewollten Auseinandersetzung mit einem Lerngegenstand entsteht eine positive Lernmotivation und die motivationale Grundlage für Bildungsprozesse. Nach Auffassung moderner Interessentheorien ist dieses zugleich eine entscheidende Grundlage für das Auftreten intrinsischer Motivation (Schiefele & Schreyer, 1994). Daher wird in vielen theoretischen Ansätzen heute der Begriff „Interesse“ als Synonym für „intrinsische Motivation“ verwandt (Prenzel, Krapp & Schiefele, 1986).

Zusammenfassend lassen sich folgende Merkmale des Interesses – übertragen auf das Studium – anführen (Prenzel, Krapp & Schiefele, 1986; Schiefele, Winteler & Krapp, 1988):

- *Wertschätzung*: Das Studium erfährt eine hohe Wertschätzung, das heißt es nimmt einen hohen Rangplatz in der individuellen Wertehierarchie ein. Eine Auseinandersetzung mit dem Studienstoff ist an sich für die Person wertvoll und wird anderen Verhaltensweisen vorgezogen. Entsprechend wird mehr Zeit für eine Auseinandersetzung mit dem Gegenstand verbracht (Rolfs, 2001).
- Damit eng verwoben ist die so genannte *Selbstintentionalität*: Die mit dem Studium verbundenen Tätigkeiten werden um ihrer selbst willen – also ohne äußeren Anlass – ausgeführt. Dabei motiviert die Auseinandersetzung selber, nicht aber das Erreichen bestimmter Ziele (H. Schiefele, Prenzel, Krapp, Heiland & Kasten, 1983, zitiert nach Krapp, 1992). Ebenso besteht Freiwilligkeit in der Handlungsausübung (Deci & Ryan, 1993).
- *Emotional*: Die Beschäftigung mit den Studienthemen ist mit positiven emotionalen Gefühlszuständen verknüpft. Bei einer Auseinandersetzung mit den Stoffen kommt es unter anderem zu einem *flow*-Erleben. Dieses Erleben umschreibt Csikszentmihalyi (1985) als „Verschmelzen von Handlung und Bewusstsein“ (S. 61). Umfassender charakterisiert Rheinberg diesen Zustand als „selbstreflexionsfreie[s], gänzliche[s] Aufgehen in einer glatt laufenden Tätigkeit, bei der man trotz voller Kapazitätsauslastung das Gefühl hat, den Geschehensablauf noch voll unter Kontrolle zu haben“ (2006, S. 345). Darüber hinaus wird schon alleine der Gedanke an den Stoff von angenehmen Gefühlen begleitet.
- *Einfluss auf verwandte Lernstrategien*: Sowohl das globale als auch das spezifische Studieninteresse für ein bestimmtes Thema unterstützt elaborierte Lernstrategien, so dass eine tiefere kognitive Wissensverarbeitung angestrebt wird (Wild, Krapp & Winteler, 1992). Dieser so genannte *deep approach* ist charakterisiert durch eine breit gefächerte Lektüre während der Prüfungsvorbereitung und den Versuch, die jeweiligen neuen Informationen mit früherem relevanten Wissen in Beziehung zu setzen (Wild, 2001). Es steht im Gegensatz zum oberflächlichen Lernen (*surface approach*), bei welchem der Lernstoff lediglich auf das Notwendige beschränkt wird. Dabei steht das Auswendiglernen von Fakten im Mittelpunkt, nicht das Verstehen des Stoffes.

- *Kognitiv*: Das Interesse zeigt – indirekt durch die andere Lernstrategie bestimmt – ferner Auswirkungen auf das gegenstandsspezifische Verständnis. Daraus resultieren differenzierte Handlungsmöglichkeiten, differenzierteres und integrierteres Wissen und eine gute Gegenstandsauffassung. Metaanalytisch zeigten sich Korrelationen von $r = 0,30$ zwischen dem Interesse und dem Wissen über bestimmte Sachverhalte (Schiefele, Krapp & Schreyer, 1993), wobei aber die Wirkungsrichtung unklar bleibt.

Die Forschung der letzten Jahre zum beruflichen Interesse hat – vergleichbar zu den Intelligenzstrukturmodellen – zu hierarchischen Strukturmodellen des Interesses geführt. Ziel dieser Modelle ist es, die Basis für die *measured interests* zu legen, dabei Unschlüssige bei der Berufsentscheidung oder Studienfachwahl zu unterstützen und die oben beschriebenen positiven Effekte zu erzielen. Prediger (1982) konnte nachweisen, dass in zahlreichen Interessendaten ein genereller Interessenfaktor vorherrscht – vergleichbar zur *General Mental Ability* in der Intelligenzforschung (siehe Abschnitt 3.3.1). Auch unabhängig von den differenzierten Interessen scheint es Personen zu geben, die zahlreiche Tätigkeiten gerne ausführen und wiederum andere, die nur wenige präferieren und diese um so häufiger ausführen. Abseits von diesem generellen Interessenfaktor, der natürlich für die weitere Beratung nur von geringerer Bedeutung ist, stehen in der Literatur zahlreiche unterschiedliche Interessenstrukturmodelle mit ganz unterschiedlichen Faktorenanzahlen zur Diskussion. So postuliert Prediger (1982) zwei weitere Faktoren, die das Interesse an der Auseinandersetzung mit *Personen* („Interpersonal tasks such as caring for, persuading, entertaining, or directing others, including animals treated as if they were human“, S. 261) versus *Dingen* („Nonpersonal tasks involving machines, materials, tools, biological mechanisms, and so forth“, S. 261) und *Daten* („Impersonal tasks involving facts, records, files, numbers, and systematic procedures for assisting goods/services consumption by people“, S. 261) versus *Ideen* („Intrapersonal tasks involving abstractions, theories, knowledge, insights, and new ways of expressing something, for example, with words, equations, or music“, S. 261) beschreiben.

Die einflussreichste Theorie des beruflichen Interesses (Gottfredson, 1999; Holling, Lücken, Preckel & Stotz, 2000) stammt von John L. Holland (1997). Neben dem Ansatz von Parson (siehe Abschnitt 2.2.1) und dem Werk Murrays, der bereits 1938 die Annahme vertrat, dass Verhalten sowohl von der Persönlichkeit als auch von der Umwelt abhängt, ist Hollands Theorie insbesondere von dem Versuch von Guilford, Christensen, Bond und Sutton (1954) inspiriert worden, die Faktorenstruktur des Interesses auf empirischem Weg genauer zu spezifizieren. Anhand einer Faktorenanalyse von über 1 000 Tätigkeitsformulierungen als Itemmaterial ließen sich dabei sechs Faktoren finden (S. 28), die bis heute die sechs Faktoren der Theorie Hollands bilden (siehe Tabelle 3.1):

- *mechanical interest* wurde in der Theorie Hollands zu *realistic*
- *scientific interest* wurde in der Theorie Hollands zu *investigative*

- *social welfare interest* wurde in der Theorie Hollands zu *social*
- *aesthetic expression interest* wurde in der Theorie Hollands zu *artistic*
- *clerical interest* wurde in der Theorie Hollands zu *conventional*
- *business interest* wurde in der Theorie Hollands zu *enterprising*

Die Grundlage der Theorie Hollands besteht in den folgenden zentralen Aussagen zum Individuum, zur Umwelt und zur Kongruenz zwischen Individuum und Umwelt (Zimmerhofer, 2003, S. 45-47):

- Aufgrund sozialer, kultureller und biologischer Kräfte bilden sich bei Heranwachsenden Präferenzen für bestimmte Tätigkeiten und dann Interessendispositionen, so dass eine Person entsprechend wahrnimmt, denkt und handelt.
- *Personen*
 - Menschen westlicher Kulturräume lassen sich anhand von sechs Interessentypen (als zeit- und situationsstabile Persönlichkeitsmerkmale verstanden) beschreiben (siehe Tabelle 3.1 und Abbildung 3.2).
 - Die Wahl eines Berufes ist somit Ausdruck der Persönlichkeit.
 - Personen mit gleichen beruflichen Interessen haben ähnliche Persönlichkeiten und ähnliche Entwicklungen durchlaufen.
 - Der persönliche Interessen-Code – im Folgenden *Three-Letter-Code* genannt – wird gebildet durch jeweils drei Interessendimensionen der Theorie Hollands, bei der die betrachtete Person die höchsten Skalenscores zeigt.

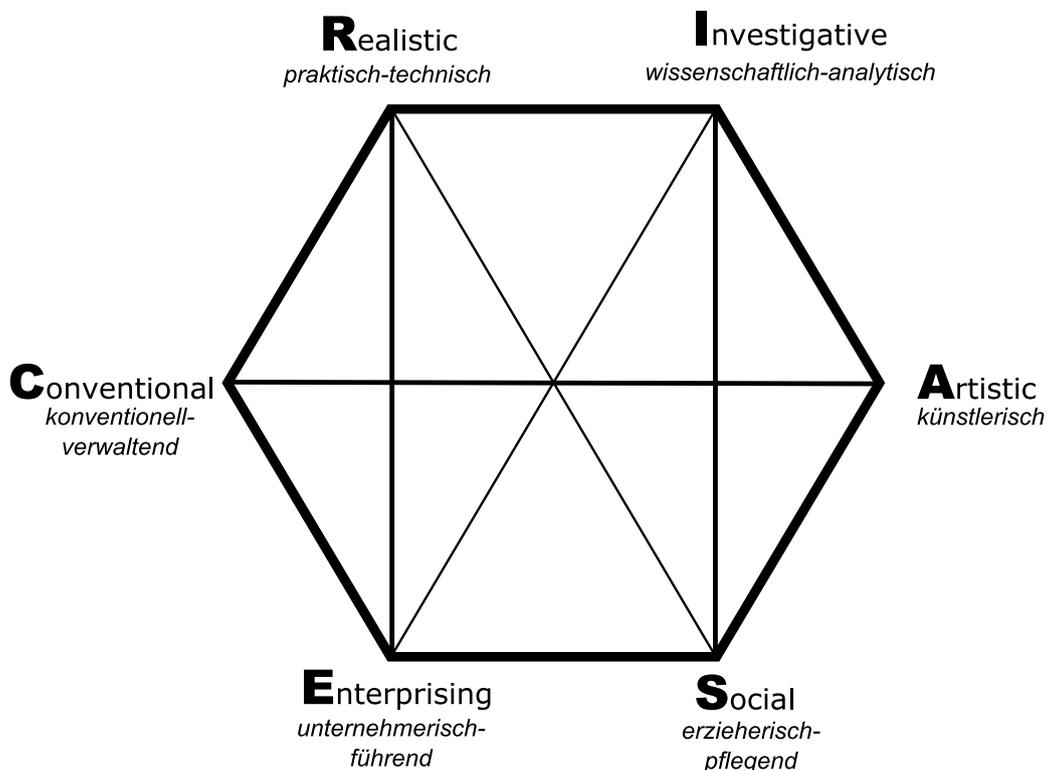


Abbildung 3.2: Hexagon der beruflichen Interessen nach Holland (1997). Die Dicke der Verbindungslinien symbolisiert die postulierte Stärke des statistischen Zusammenhangs.

- *Umwelt*
 - Analog wird dazu die Umwelt ebenfalls anhand dieser Dimensionen beschrieben.
 - Dabei werden die physikalischen Bedingungen, die Tätigkeitsstruktur (Berufskongruenz) und die beruflichen Interessen der Angehörigen (Umweltkongruenz) berücksichtigt.
 - Vergleichbar zum Three-Letter-Code der Person kann auch die Umwelt anhand von drei Dimensionen beschrieben werden. Meist werden Experten zur Festlegung der Umweltbeschreibung befragt.
- *Personen-Umwelt-Interaktion*
 - Menschen suchen nach einer Umwelt, die ihnen erlaubt, ihre Fertigkeiten und Fähigkeiten einzusetzen und ihre Einstellungen und Werte auszuleben.
 - Aus der besonderen Passung zwischen Interesse und Umwelttyp (Personen-Umwelt-Kongruenz) leitet sich das Erleben und Verhalten von Personen ab.

Die Theorie Hollands erwies sich für die Forschung sowie für den praktischen Einsatz in Berufsberatungen in der westlichen Welt als sehr fruchtbar. Zahlreiche Publikationen sind in den letzten Jahrzehnten gerade über die faktorielle Validität (zum Beispiel Rounds, Tracey & Hubert, 1992; Tracey & Rounds, 1993), die Kriteriumsvalidität (Assouline & Meir, 1987; Spokane, 1985; Spokane, Meir & Catalano, 2000; Tranberg, Slane & Ekeberg, 1993), aber auch über die Generalisierbarkeit der Interessenstruktur (zum Beispiel für den Einsatz auf dem asiatischen Kontinent, Long & Tracey, 2006; Rounds & Tracey, 1996; Tracey, Watanabe & Schneider, 1997) und zu Geschlechts- beziehungsweise sozioökonomischen Unterschieden (Ryan, Tracey & Rounds, 1996) veröffentlicht worden.

Gerade die faktorielle Validität stand häufig zur Debatte; zahlreiche Studien verglichen die Faktorenstruktur mit denen der alternativen Modelle von Prediger oder Roe. Das von Holland beschriebene *Calculus*¹⁶- (letzteres wird teilweise auch *circular order model* genannt) beziehungsweise *Circumplex*¹⁷-Modell lässt sich anhand von drei unterschiedlichen Methoden überprüfen. Die Verwendung der konfirmatorischen Faktorenanalyse, der Multidimensionalen Skalierung und der Randomization Test of Hypothesized Order Relationships (Hubert & Arabia, 1987) sind hier Methoden der Wahl. Dabei zeigt sich, dass die von Holland postulierten Faktorbeziehungen zu einem großen Anteil auch empirisch bestehen. Rounds, Tracey und Hubert (1992) konnten in einer Stichprobe 62 (weibliche Personen) beziehungsweise 69 (männliche

¹⁶ Zentral bei der Calculus-Annahme ist, dass die einzelnen Interessendimensionen in einem Hexagon angeordnet sind und die Distanzen zwischen diesen im Hexagon umgekehrt proportional zum theoretischen Zusammenhang sind. Dabei sind insgesamt 72 Interkorrelationen zu betrachten, nämlich die direkt benachbarten, die gegenüberliegenden und die diagonal gegenüberliegenden Interessendimensionen – siehe auch Abbildung 3.2.

¹⁷ Die Calculus-Annahme wird hierbei um 33 Vorhersagen erweitert. Ergänzt wird die Ordnung der jeweiligen Korrelationen innerhalb der drei unterschiedlichen Zusammenhangsgruppen. Nach Holland (1997) sollen die Interkorrelationen der benachbarten, der gegenüberliegenden und der nicht direkt benachbarten Dimensionen in etwa gleich groß sein. Insgesamt werden also statt 72 105 Relationen postuliert.

Tabelle 3.1
Kurze Beschreibung der Interessen-Dimensionen nach Holland (1997)

	Realistic	Investigative	Artistic	Social	Enterprising	Conventional
	Praktisch-technisch	Intellektuell-forschend	Künstlerisch-sprachlich	Sozial	Unternehmerisch	Konventionell
Beschreibung der Tätigkeiten	Tätigkeiten, die Kraft, Koordination und Handgeschicklichkeit erfordern und zu konkreten, sichtbaren Ergebnissen führen	Tätigkeiten, bei denen das Ergebnis anhand systematischer Beobachtung und Forschung im Mittelpunkt steht	Offene, unstrukturierte Tätigkeiten, die künstlerisch kreative Produkte ermöglichen	Tätigkeiten in Interaktion mit Menschen, wie Lehren, Versorgen oder Beraten	Tätigkeiten, bei denen Personen beeinflusst, geführt und/oder manipuliert werden	Tätigkeiten, bei denen strukturiert, standardisiert und regelhaft vorgegangen wird
Unterstützt werden seitens der Umwelt	Konformes Verhalten, praktische Leistungen	Hohes Beharungsvermögen und schlussendliches Lösen von Problemen	Einfallstreichtum in Literatur, Kunst oder Musik	Empathie, humane Einstellung, Geselligkeit und Wohlwollen	Streben nach finanziellen und/oder materiellen Dingen, ferner Dominanz und Selbstsicherheit	Organisationsfähigkeit, Konformität, Verlässlichkeit
Berufe (beispielhaft)	Schlosser, Tischler, Maschinenbauer, Bauingenieur	Mathematiker, Informatiker	Musiker, Schauspieler, Schriftsteller, Designer	Therapeut, Krankenpfleger, Pädagoge	Rechtsanwalt, Vertreter, Geschäftsführer	Verwaltungsbeamter, Buchhalter

Personen) von insgesamt 72 Prädiktionen des Modells ansatzweise nachweisen. Im direkten Vergleich zu anderen Interessenmodellen (zum Beispiel dem von Gati, siehe Tracey & Rounds, 1993) zeigt sich bei Hollands Modell eine größere Übereinstimmung zwischen den Vorhersagen des Modells und der Empirie.

Um den *Person-Vocation-Fit* (siehe Abschnitt 2.2.1) zu steigern, ist anzustreben, dass eine Berufswahl oder auch Studienfachwahl auf Basis der Passung zwischen der eigenen Persönlichkeit beziehungsweise dem Selbstkonzept und der Anforderungen des Studienfaches geschieht (Holland, 1997). Auch wenn über den zentralen Bestandteil der Theorie nach Holland lebhaft debattiert wird (Dawis, 2000; Rounds, McKenna, Hubert & Day, 2000; Tinsley, 2000a; Tinsley, 2000b; Tracey, Darcy & Kovalski, 2000), konnten zahlreiche Studien einen positiven Zusammenhang zwischen der Interessenkongruenz nach Holland und diversen Kriterien aufzeigen: Nachweisbar waren in diesem Sinne großenteils in Metaanalysen sowohl ein Zusammenhang mit der Zufriedenheit am Arbeitsplatz (Assouline & Meir, 1987; Spokane, 1985; Spokane, Meir & Catalano, 2000; Tranberg, Slane & Ekeberg, 1993) als auch mit der Studienfachwahl (Rolfs, 2001; Spies, Westermann, Heise & Hagen, 1998), eine bessere Arbeits- (Tsabari, Tziner & Meir, 2005) beziehungsweise Studienleistung (Bergmann, 1992; Rolfs, 2001), eine höhere Stabilität der Studienfachwahl (Bergmann, 1992; Brandstätter, Farthofer & Grillich, 2001; Brandstätter, Grillich & Farthofer, 2002; Rolfs, 2001) und eine bessere Tiefenverarbeitungsstrategie bei der Prüfungsvorbereitung (Rolfs & Schuler, 2002b).

Auch wenn Interesse unter anderem die Zufriedenheit mit einer Umwelt (hier ein bestimmtes Studienfach) voraussagen kann, zeigen sich in Studien im Vergleich zu kognitiven Verfahren nur geringe Korrelationen: Optimistisch sprach Spokane (1985) von einem „magic .30 correlation plateau“ (S. 335). Metaanalysen zeigen hingegen lediglich eine mittlere, unkorrigierte Korrelation von $r = 0,174$ (Tranberg, Slane & Ekeberg, 1993), von $r = 0,166$ ¹⁸ (Tsabari, Tziner & Meir, 2005) beziehungsweise von $r = 0,210$ (unkorrigiert, Assouline & Meir, 1987). Diese geringen statistischen Zusammenhänge lassen sich aber wohl zu einem größeren Anteil auf wenig geeignete Interessenkongruenz-Indizes zurückführen (Rolfs, 2001). Im studentischen Umfeld konnte Bergmann (1992) aufzeigen, dass Maturanten, die ein Studium passend zu ihren gemessenen Interessen gewählt haben, zufriedener und weniger belastet waren. Bezüglich der Ingenieurwissenschaften waren der Studienerfolg (Selbsteinschätzung $r = 0,24$), die Studienzufriedenheit ($0,41$) und die Wechselbereitschaft ($-0,34$) prognostizierbar. Zudem zeigten sie einen stabileren Studienverlauf (Brandstätter, Farthofer & Grillich, 2001). Rolfs (2001) wies nach, dass gerade bei Verwendung effizienter Kongruenzindizes mit einer gesteigerten Interessenkongruenz der Lernaufwand höher ($r = 0,23$), das subjektive Wohlbefinden ($r = 0,57$) besser und die Gesamtbewertung des Studiums ($0,53$) positiver ausfielen. Zudem war die Absicht, das Studium abzubrechen ($-0,35$) geringer. Dabei zeigte sich für das Erleben des Studiums eine hohe inkre-

¹⁸ Die Ergebnisse sind doppelt minderungskorrigiert mit Berücksichtigung der Varianzeinschränkung.

mentelle Validität auch über die Persönlichkeitsdimensionen Extraversion, Gewissenhaftigkeit und Verträglichkeit hinaus (Rolfs & Schuler, 2002a).

3.4.4 Extrinsische Lernmotivation

Im vorigen Abschnitt wurde der Begriff des Interesses und die damit verbundene *intrinsische Lernmotivation* eingeführt. Diese Form der Lernmotivation wird immer dann wirksam, wenn „das Motivationsziel subjektiv in der jeweiligen Handlung besteht, die Handlung also subjektiv ihr eigenes Ziel darstellt, indem sie in entsprechenden Handlungswünschen als Ziel repräsentiert ist“ (Pekrun, 1988, S. 195). Ein Beispiel: Ein Student befasst sich gerne in seinem Studium mit dem Themengebiet der Statistik, so dass er ohne äußeren Anreiz neuen Stoff durcharbeitet, vielleicht gar schon vorarbeitet. Er vollzieht die Handlung zudem freiwillig (Deci & Ryan, 1993). Im Rahmen der Studienberatung und Berufsberatung wird im Allgemeinen angestrebt, Berufe und Studiengänge zu finden, die für den Ratsuchenden interessant sind, so dass der oben beschriebene Umstand möglichst häufig eintritt.

Im Allgemeinen wird dieser Form die *extrinsische* Lernmotivation gegenüber gestellt. Extrinsisch soll dabei „äußerlich“ oder „nicht wirklich dazugehörend“ heißen. Was aber die einzelne Bezugsgröße für das „Innen“ und „Außen“ ist, schwankt von Betrachter zu Betrachter, von Autor zu Autor (Rheinberg, 1995). Entsprechend herrscht in der Literatur kein Konsens darüber, wie beide Motivationsquellen exakt definiert und insbesondere auch deutlich unterschieden werden können. Zudem wird die extrinsische Lernmotivation häufig nur indirekt über das Komplement der intrinsischen Motivation beschrieben. Dabei wird definiert, dass die Handlung entsprechend nicht um ihrer selbst Willen vollzogen wird. Der Grund für eine Auseinandersetzung mit einem Gegenstand – zum Beispiel für eine Lernhandlung – liegt also bei äußeren Handlungsveranlassungen; sie wird zum Beispiel hervorgerufen durch Lob oder der Vermeidung von Bestrafung. Somit werden die Handlungen mit instrumentellen Absichten begonnen, welches eine von der Handlung selbst separierbare Konsequenz darstellt (Deci & Ryan, 1993). Es wird versucht, bestimmte Ergebnisse zu erreichen (zum Beispiel ein gutes Prüfungsergebnis) beziehungsweise eine bestimmte Konsequenz zu vermeiden (zum Beispiel den Hohn und Spott der Kommilitonen).

Lange Zeit galten die *intrinsische* und *extrinsische* Motivation nicht nur als direkte Gegenpole, sondern auch als positive und negative Quelle der Motivation. Die letztere Ansicht wurde gerade nach den Studien von Deci (1971) populär, da dort extrinsische Motivationsquellen, wie Bezahlung, die intrinsischen korrumpierten. U. Schiefele und Schreyer (1994) sprechen dazu analog „von einem gegeneinander arbeiten“ der Faktoren (S. 11). Ist die intrinsische Lernmotivation eher mit einem gründlichen Lernen verbunden, so zeigt die extrinsische Lernmotivation

einen signifikanten Zusammenhang (aber lediglich mit 5 % Varianzaufklärung) mit den Oberflächenstrategien des Lernens (Schiefele & Schreyer, 1994).

Die Korruption der extrinsischen Motivation ist aber nicht allgemeiner Konsens. So sieht Pekrun (1988) eher eine additive Wirkung extrinsischer und intrinsischer Motivation. Unterstützt wird diese Ansicht durch die Metaanalysen von Cameron und Pierce (1994) beziehungsweise Eisenberger und Cameron (1996), die keine generellen negativen Effekte von extrinsischer Verstärkung aufzeigen. Aber auch Deci selbst geht heute entsprechend seiner Selbstbestimmungstheorie (Deci & Ryan, 1985) nicht mehr von einem einfachen gegensätzlichen Effekt der extrinsischen und intrinsischen Quellen aus. Vielmehr soll der resultierende Effekt von der Interpretation des Individuums abhängen, was die unterschiedlichen Ergebnisse der Studien begründen könnte: Entweder wird die Belohnung dergestalt interpretiert, dass sie das eigene Verhalten determiniert oder aber auch, dass die Belohnung nur ein Indikator der eigenen Kompetenz ist (Deci, Koestner & Ryan, 1999). Dabei sind intrinsisch und extrinsisch kein Gegensatzpaar, sondern vielmehr ein Kontinuum, das sich durch den Grad der Selbstbestimmung unterscheidet (Deci & Ryan, 1993; Deci & Ryan, 1985): Der tüchtige Student, der eine gute Note erzielen möchte, um ein klar umrissenes Oberziel (ein langfristiges, übergeordnetes Ziel) zu erreichen, ist zwar nicht intrinsisch, sondern vielmehr extrinsisch motiviert – dafür aber ebenfalls hoch selbstbestimmt. Dieser Fall wird von den Autoren *integrierte Regulation* genannt, die der intrinsischen Motivation auch in den positiven Auswirkungen sehr ähnlich ist. Selbstbestimmte Formen der Motivation führen zu insgesamt besseren Lernergebnissen, auch bei einer statistischen Kontrolle des Fähigkeitsniveaus (Black & Deci, 2000). Die entsprechend motivierten Personen schneiden insbesondere in Bezug auf konzeptionelles Lernen besser ab und streben höherwertige Schulabschlüsse an. Weiterhin verwenden sie günstigere Copingstrategien nach Misserfolg (Krapp, 2003).

Da in einem Studium nicht nur für die Studierenden interessante Dinge gelernt werden müssen, kann der Student sich nicht ausschließlich auf intrinsische Motivationsquellen verlassen. Nach Heckhausen (1989) ist die extrinsische Motivation durch die folgenden Ziele charakterisiert: So kann eine Person dadurch motiviert werden, eine positive Bewertung durch sich selbst zu erreichen. Sie möchte also stolz auf die erbrachten Leistungen sein können. Ein weiteres Element widmet sich der Bewertung durch andere Personen, durch Lehrer, Dozenten oder Kommilitonen. Eine anvisierte positive Fremdbewertung kann zum Beispiel durch ein Lob ausgedrückt werden. Eine weitere Quelle für eine extrinsische Lernmotivation ist die Annäherung an Oberziele. So könnte eine Person motiviert sein, den Stoff für eine wichtige Abschlussprüfung gut zu erarbeiten, um mit einer guten Examensnote später einen guten Start ins Berufsleben schaffen zu können. Die betrachtete Lernhandlung ist dabei also ein Schritt von mehreren zur Erreichung eines weiter reichenden Ziels, nämlich den Berufseinstieg. U. Schiefele und Urhahne (2000) unterscheiden dabei vergleichbar leistungs- (gute Studienleistung), wettbewerbs- (besser sein als andere) und berufsorientierte (Erreichen beruflicher Ziele) extrinsische Lernmotivatio-

nen. Dabei spielen bei der Leistungs- und Wettbewerbsorientierung sowohl Selbst- als auch Fremdbewertungen eine Rolle. Die Berufsorientierung lässt sich hingegen klar der Orientierung an ein Oberziel nach Heckhausen zuordnen.

3.4.5 Selbstwirksamkeit

Menschen schätzen vor der Ausführung von bestimmten Handlungen ab, wie wahrscheinlich bestimmte Ergebnisse durch Handlungen erzielt werden können (siehe Erwartung-mal-Wert-Modelle, vgl. Rheinberg, 1995). Sie antizipieren also die Zukunft und entscheiden auf der Basis der erwarteten Ergebnisse (und deren Bewertung) entsprechend, ob sie eine bestimmte Verhaltensweise ausüben oder nicht. Diesen Entscheidungen liegen zwei unterschiedliche Formen der Erwartung über die Zukunft zu Grunde: die Wirksamkeitserwartung und die Ergebniserwartung (siehe Grafik 3.3).

Übertragen auf das pädagogische Umfeld lässt sich der Prozess wie folgt an einem Beispiel spezifizieren: Eine Person steht kurz vor einer Prüfung und müsste sich eigentlich mit einem bestimmten Lehrbuch vorbereiten. Zuerst schätzt die Person als *Wirksamkeitserwartung* ab, ob sie überhaupt in der Lage ist, das entsprechende Lernverhalten – hier also das Durcharbeiten des Buches – zu zeigen. Geht die Person zum Beispiel davon aus, dass sie das Buch nicht komplett durcharbeiten kann, da es zu schwer verständlich verfasst ist, wird sie dieses Lernverhalten nicht weiter in Betracht ziehen. Wird jedoch die Erwartung als hoch eingeschätzt, steigt auch die Wahrscheinlichkeit, dass das Verhalten gezeigt wird. Menschen schätzen also auf der Basis früherer Erfahrungen ein, inwieweit die entsprechende Fähigkeit für die betrachtete Verhaltensweise überhaupt vorhanden ist. Bandura nennt dies selbst die Erwartung „to organize and execute given types of performance“ (Bandura, 1977b, S. 21). Wichtig dabei ist, dass dies eine rein subjektive, unter Umständen verzerrte Einschätzung der eigenen Selbstwirksamkeit ist, die nicht unbedingt mit den objektiven Fähigkeiten zu tun haben muss.

Zentral sind aber auch die *Ergebniserwartungen* (response outcome expectations), die unabhängig von den bisher beschriebenen Selbstwirksamkeitserwartungen sind: Wird im obigen Bei-

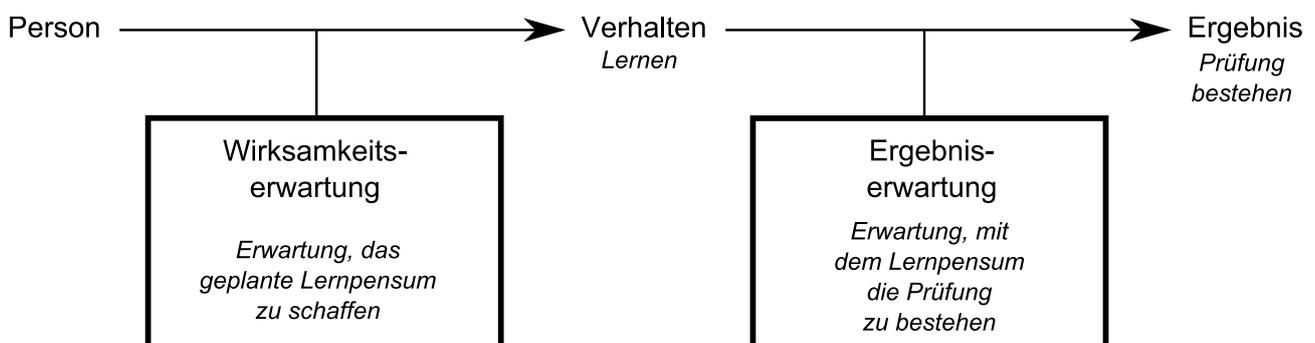


Abbildung 3.3: Veranschaulichung des Zusammenspiels von Wirksamkeitserwartungen und Ergebniserwartungen (modifiziert nach Bandura, 1977a)

spiel angenommen, dass auch das Durcharbeiten und Verstehen des Buches nicht zu einer guten Prüfung führt (vielleicht, weil der Prüfer als unfair eingeschätzt wird), sinkt die Wahrscheinlichkeit, dass Verhalten auch zu zeigen. Somit greifen hier bestimmte erlernte Kontingenzen von Verhalten und Verhaltensfolgen. Bandura selbst umschreibt diese Erwartung mit „judgement of the likely consequence such performance will produce“ (Bandura, 1977b, S. 21).

Seit den ersten Veröffentlichungen von Bandura zu diesem Thema vor etwa 30 Jahren (Bandura, 1977a; Bandura, 1977b) konnte die Bedeutung der jeweiligen spezifischen Selbstwirksamkeitserwartungen für viele Lebensbereiche nachgewiesen werden: Zusammenhänge mit klinischen Forschungsgegenständen wie Phobien (Bandura, 1983), Drogenabhängigkeiten (Ilgen, McKellar & Tiet, 2005), Ernährungsgewohnheiten (Steptoe, Perkins-Porras, Rink, Hilton & Cappuccio, 2004), Erfolge der Rehabilitation (Bray & Cowan, 2004; Middleton, Tate & Geraghty, 2003) oder Stress (Jex, Bliese, Buzzell & Primeau, 2001) wurden berichtet. Judge und Bono (2001) wiesen einen Einfluss der Selbstwirksamkeit auf die Arbeitszufriedenheit ($r = 0,45$, korrigiert¹⁹) und auf die Arbeitsleistung ($r = 0,23$) nach. Auch im Karriereverlauf, also bei den einzelnen Entscheidungen in Berufs- und Studienwahl, spielen Selbsterwartungen bezüglich einzelner Berufe oder Studienfächer eine große Rolle: Die entsprechende Selbstwirksamkeitserwartung zeigt allgemein einen Zusammenhang mit der Anzahl, Breite, aber auch mit der Art bedachter Berufsalternativen (Betz & Hackett, 1981).

Zahlreiche Studien befassen sich ferner mit dem Zusammenhang von Selbstwirksamkeitserwartungen und der Lernmotivation (Krapp & Ryan, 2002; Zimmerman, 2000) und dem resultierenden Schul- beziehungsweise Studienerfolg (Chemers, Hu & Garcia, 2001; Multon, Brown & Lent, 1991; Robbins et al., 2004; Urhahne, 1997; Wolfe & Johnson, 1995). Dabei wird nach Bandura (1997) davon ausgegangen, dass Selbstwirksamkeit eine positive Korrelation mit den zentralen Indikatoren für erhöhte Lernmotivation erreicht. Ein Ansteigen der Lernaktivitäten, Lernpersistenz, Anstrengung und emotionalen Reaktionen wird somit postuliert (Zimmerman, 2000). Multon, Brown und Lent (1991) veröffentlichten Ergebnisse einer Metaanalyse, die einen Zusammenhang der Selbstwirksamkeit mit der Persistenz im Studium ($r = 0,34$) und dem Studienerfolg (gemessen anhand von Studiennoten, Testergebnissen, Selbstauskünften und Arbeitsproben) ($r = 0,38$) zeigen. Etwa zwei Drittel der betrachteten Probanden waren aber noch nicht im Tertiärbereich, hatten also noch kein Hochschulstudium begonnen. Kahn und Nauta (2001) betrachteten in Erweiterung dazu in ihrer Studie Universitätsstudenten, also Studierende im Tertiärbereich des US-amerikanischen Bildungssystems. Dabei zeigten sich weniger klare Ergebnisse: Zwar ließ sich nachweisen, dass die Selbstwirksamkeitserwartungen des zweiten Semesters ein signifikanter Prädiktor für die Aufnahme des zweiten Jahres (sophomore year) ist. Sie konnten jedoch keinen Einfluss der Selbstwirksamkeitserwartungen vor dem College auf die Persistenz bis zum zweiten Jahr hin belegen. Gore (2006) wies Vergleichbares nach: Die akademische Selbstwirksamkeit – erhoben vor dem Studium – erwies sich als schwacher Prädiktor

¹⁹ Die Ergebnisse sind doppelt minderungskorrigiert.

für den späteren Studienerfolg. Als bedeutender erwies sich der statistische Zusammenhang zwischen der akademischen Selbstwirksamkeit am Ende des ersten Semesters und den Studiennoten des zweiten ($r = 0,35$) beziehungsweise dritten Semesters ($r = 0,21$). Anscheinend war die Selbstwirksamkeit erst dann prädiktiv, wenn die Wirksamkeitserwartungen bereits auf der Basis von Erfahrungen im Studium entstanden sind. Moderiert wird die Prädiktionsleistung wahrscheinlich durch den Umstand, inwieweit Schüler bereits Erfahrungen mit der akademischen Selbstwirksamkeit (also mit Lern- und Prüfungsverhalten) gemacht haben.

In einer weiteren Metaanalyse untersuchten Robbins et al. (2004) unter Berücksichtigung von 109 Studien den Einfluss nicht-kognitiver Variablen auf die kumulierten Studiennoten (*cumulative GPA*) und auf die Persistenz. Dabei zeigte sich die akademische Selbstwirksamkeit als bester Prädiktor ($r = 0,496^{20}$) für die Studiennoten und als zweitbesten ($r = 0,359^{20}$) für die Persistenz. In einer rein longitudinal angelegten Studie (Chemers, Hu & Garcia, 2001) verlief bei einem Pfadmodell der Einfluss der Selbstwirksamkeit auf den Studienerfolg sowohl direkt als auch indirekt, und zwar vermittelt über Bewältigungsstrategien, Stress, Gesundheit, generelle Zufriedenheit und Commitment. Auch mit experimentell induzierten Selbstwirksamkeitserwartungen (Bouffard-Bouchard, 1990; Cervone & Peake, 1986) ließ sich zeigen, dass stärker wirksame Probanden eine höhere Strategieflexibilität für Problemlösungen verwandten und entsprechend auch eine bessere Leistung zeigten. In einer Studie von DeWitz und Walsh (2002) ließ sich die Selbstwirksamkeit auch für die Prognose der Studienzufriedenheit nutzen: So klärte ein Fragebogen zur akademischen Selbstwirksamkeit alleine 24,1 % Varianz auf. Weder die soziale noch die generelle Selbstwirksamkeit konnte die Prädiktion der Studienzufriedenheit signifikant verbessern.

Neben Studienabbruch, Studiennoten und Studienzufriedenheit lässt sich auch die Procrastination, also das Aufschieben von Prüfungen, anhand von Verfahren zur Selbstwirksamkeit vorhersagen. Helmke und Schrader (2000) konnten über eine Fragebogenerhebung eine Korrelation zwischen der generellen Selbstwirksamkeit und der Procrastination von $r = -0,38$ (state) beziehungsweise $-0,39$ (trait) nachweisen. Mangelnde Selbstwirksamkeit in Bezug auf die eigene Leistungsfähigkeit scheint dafür mitverantwortlich zu sein, dass Lernhandlungen während der Prüfungsvorbereitung nicht oder erst zu spät begonnen werden. Die Lernenden gehen in diesen Fällen davon aus, dass sie die für ein Bestehen nötigen Fähigkeiten nicht besitzen.

3.4.6 Handlungs- versus Lageorientierung

Im Kapitel zum Interesse beziehungsweise zur intrinsischen Motivation (siehe Abschnitte 3.4.3) zeigte sich, dass sich Personen mit bestimmten Dingen gerne auseinander setzen und dies auch einen Einfluss auf die Beständigkeit und Ergebnis der Handlungen haben kann. Doch im

²⁰ Die Ergebnisse sind doppelt minderkorrigiert.

Leben eines Studenten geht es nicht nur darum, Dinge zu vollziehen, die man gerne ausführt. Die extrinsische Lernmotivation (siehe Abschnitt 3.4.4) kann motivieren, wenn Stoffe erarbeitet werden müssen, die zwar nicht interessant erscheinen, aber ein Instrument für ein mittelfristiges Ziel darstellen. Häufig sind Entscheidungen getroffen, mit dem Lernverhalten zu beginnen, nur muss dieser Plan noch umgesetzt werden (siehe Abschnitt 3.4.2): Sie müssen sich zum Beispiel auf Prüfungen vorbereiten, während Freunde das schöne Wetter genießen, oder Referate halten, obwohl sie die Situation, vorne vor einem Publikum stehen zu müssen, überhaupt nicht schätzen. Im Folgenden wird der Fokus daher auf der Realisierung von Handlungen liegen, deren Ausführung für den Studenten eher einen abschreckenden Charakter haben. Für all diese Verhaltensweisen müssen die Personen sich umgangssprachlich „zusammenreißen“ oder „den inneren Schweinehund überwinden“: Nach der Planungsphase sind eben innere Widerstände zu überwinden, um die Tätigkeit zu initiieren und schlussendlich ein bestimmtes Ziel zu erreichen. Schon früh in der psychologischen Forschung wurden diese Umstände mit dem Begriff *Wille* umschrieben.

Ach (1935) hat bereits vor mehr als 100 Jahren den Begriff *Wille* in die Forschung eingeführt. Er untersuchte Verhaltensweisen von Personen, die gegen eine fest eingeschlifene Gewohnheit handeln mussten. Ach gab Probanden die Aufgabe, bedeutungsfreie aber sich reimende Silbenpaare zu lernen. In der zweiten Phase wurde eine zusätzliche Operation verlangt: Zum Beispiel wurde die erste Silbe eines Paares genannt, jetzt bestand aber die Aufgabe darin, als zweites Silbenpaar die erste Silbe von hinten zu lesen. Somit mussten sich neuartige Handlungsweisen (die neue Sortierung der Silben) gegen einen automatisierten Ablauf durchsetzen. Über die Messung von Reaktionszeiten, Fehlerhäufigkeiten oder Introspektion wollte Ach Rückschlüsse auf die Willensstärke vornehmen. Ach postulierte entsprechend, dass Willensprozesse immer dann vonnöten sind, „wenn der Mensch Widerstände äußerer oder innerer Art, Hemmungen, Schwierigkeiten oder dergleichen zu überwinden hat, die sich seinem Handeln entgegenstellen“ (Ach, 1935, S. 196). Nach den Arbeiten von Ach geriet die Forschung, die sich nicht mit der Motivation, sondern mit dem Willen beschäftigt, in Vergessenheit. Die Ansicht setzte sich durch, dass die Motivationsprozesse wichtiger für die Handlungssteuerung seien (Rheinberg, 1995). Erst in den letzten Jahren ist wieder verstärkt im Bereich des Willens, nun Volition genannt, geforscht worden.

Nach Kuhl (1983) unterscheiden sich Menschen darin, wie gut es ihnen gelingt, Handlungsabsichten tatsächlich auch umzusetzen. Kuhl nennt diejenigen Menschen lageorientiert, denen es weniger gut gelingt, sich den

- a) angestrebten Zustand,
- b) den gegenwärtigen Zustand,
- c) die Diskrepanz zwischen Soll und Ist oder
- d) die beabsichtigte Handlung

bewusst zu machen.

In dieser Konzeption wird die *Handlungsorientierung bei der Tätigkeitsplanung, bei der Tätigkeitsausführung* und *nach Misserfolg* unterschieden (Kuhl, o.J.).

Fehlt eines dieser wesentlichen Elemente, kann die entsprechende Handlung nicht ausgeführt werden. Bei Studierenden können zum Beispiel die Gedanken nur um den momentanen Zustand kreisen, leider noch nicht für eine anstehende Prüfung gelernt zu haben (siehe Element b). Auch würde es nicht sinnvoll sein, ständig darüber nachzudenken, wie gut es wäre, endlich die Prüfung mit einer guten Note abschließen zu können (siehe Element a). Bei beiden unangemessenen Absichten würden die Studierenden sich mehr mit ihrer eigenen *Lage* als mit *Handlungen* (dem Lernverhalten) auseinandersetzen.

Handlungsorientierte Personen betrachten weniger die derzeitige Lage, sondern versuchen alsbald, intensiv an der Erreichung des Ziels zu arbeiten: So reagieren sie nach Problemen oder nach Misserfolgen häufig mit erhöhter Lernanstrengung, vertrauen mehr auf ihre Problemlösefähigkeit, zeigen eine stärkere Hoffnung auf Erfolg und eine geringere Versagensangst (Brunstein, 1994). Studenten mit Handlungsorientierung sind bei widrigen Umständen meist bei erhöhter Anstrengung zu effektiverem Umsetzen ihrer Lernabsichten in der Lage (Boekarts, 1994; Volet, 1997). Studenten mit Lageorientierung hingegen gelingt dies weniger. Sie empfinden entsprechend eine stärkere Frustration ($r = 0,37$) (Geyer & Lilli, 1992).

Schon in Abschnitt 3.4.2 wurde berichtet, dass die volitionalen Bestandteile u. a. die Anzahl der Prüfungen vorhersagen lassen. Entsprechend wird die Volition als wichtiger Prädiktor für das Verschleppen von Prüfungen oder Lernphasen (Helmke & Schrader, 2000) angesehen: Procrastination (das Verschleppen von Studienleistungen) korreliert mit prospektiver Lageorientierung positiv (trait: $r = 0,69$ beziehungsweise state-Procrastination: 0,59), mit Interesse (trait: $-0,42$) und Selbstwirksamkeit (trait: $-0,38$) dagegen negativ.

3.5 Zusammenfassung

In den vorigen Abschnitten wurden unterschiedliche Wege zur Prädiktion der wesentlichen Kriterien des Studienerfolgs aufgezeigt. Schulnoten als Prädiktoren sind nachweislich valide; meist stellt die Abiturdurchschnittsnote den validesten Einzelprädiktor dar. Dies gilt, obwohl die Objektivität und Reliabilität der Schulnoten – verglichen mit standardisierten Test- oder Fragebogenverfahren – durchweg geringer ausfallen. Für die Verwendung im Bereich der Studienberatung, die nicht nur die Studiennoten, sondern zum Beispiel auch die Studienzufriedenheit oder Studiendauer vorhersagen soll, ist aber die Abiturdurchschnittsnote nur eingeschränkt zu empfehlen. Eine fachspezifische Betrachtung ermöglicht die Abiturdurchschnittsnote nicht;

aber auch eine Gewichtung der Einzelnoten scheitert meist an praktischen Problemen. Ferner wird durch eine solche Gewichtung die Korrelation mit dem Studienerfolg im Vergleich zur Abiturdurchschnittsnote nicht beziehungsweise nur mäßig verbessert. Aufgrund der hohen Prädiktionsleistungen der Abiturnote muss sich aber dennoch jeder weitere Prädiktor an ihr messen; so wird meist das Inkrement über die Abiturdurchschnittsnote, also die Verbesserung der Prädiktion durch die Hinzunahme weiterer Prädiktoren, betrachtet.

Gerade Testverfahren zur Intelligenz sind effizient einsetzbar und haben sich in vielen Metaanalysen für ganz unterschiedliche Kriterien als überaus valide gezeigt. Für eine differenzierte Beratung ist eine Gewichtung der Intelligenzkomponenten möglich. Bei einer Betrachtung von Bewerbern im Bereich der Informatik und Elektrotechnik ist entsprechend eine stärkere Gewichtung der numerischen Inhalte notwendig. Bei der Entwicklung des Itemmaterials ist auch eine Kontextualisierung, also die Abstimmung des Sachverhalts der Items und des späteren Studiums, von Vorteil, um die Beratungsleistung abzurunden. Dies steigert die Akzeptanz für die Verfahren und verbessert die Informationslage der Studieninteressierten.

Für Beratungsleistungen – gerade auch für Konzeptionen mit einem Weiterbildungsaspekt – sind aber auch Persönlichkeits- und Motivationsfragebogen sinnvolle Erweiterungen, zeigen sie sich doch – je nach Konstrukt – durchaus als valide. Zahlreiche Studien zur Vorhersagekraft der Selbstwirksamkeit, der extrinsischen Lernmotivation und des Interesses wurden zusammenfassend vorgestellt. Wenn auch die Validität dieser Verfahren unter derjenigen der kognitiven Verfahren liegt, sind sie gut dazu geeignet, inkrementell die Varianzaufklärung zu verbessern, insbesondere im Bereich der Persistenz und Studienzufriedenheit. Neben dieser eher empirischen Begründung ist auch der Einfluss der Augenscheinvalidität zu beachten: Studieninteressierte erwarten geradezu einen Interessenfragebogen im Rahmen einer Studienberatung, ist doch das Interesse für ein Studienfach immer noch ausschlaggebend für die schlussendliche Fachwahl.

Kapitel 4

Ziele der vorliegenden Arbeit

Eine kritische Betrachtung des deutschen Bildungssystems stellte den Ausgangspunkt dieser Arbeit (siehe Abschnitt 1.1) dar. Deutlich wurde, dass Deutschland stärker als Nationen, die bevorzugt auf die Herstellung günstiger Massenprodukte setzen, auf einen hohen Wirkungsgrad im Bildungssystem angewiesen ist: Aufgrund der Rohstoffarmut, der hohen Lohn-, Lohnnebenkosten und der strengen Umweltauflagen etc. ist die Volkswirtschaft Deutschlands auf eine hochqualifizierte Bevölkerung angewiesen, die innovative, qualitativ hochwertige und somit auch hochpreisige Produkte für den Weltmarkt entwickelt, herstellt beziehungsweise vertreibt und wissensintensive Dienstleistungen erbringt. Gerade bei der derzeit niedrigen Geburtenrate und den hohen volkswirtschaftlichen Kosten einer Ausbildung müssen potentielle Leistungsträger möglichst frühzeitig entdeckt, effizient gefördert und gefordert werden.

Insbesondere die Erfolgsquoten eines universitären Studiums wurden als Kennwerte herangezogen, die die Effizienz des Bildungssystems abbilden können (siehe Abschnitt 1.4). Dabei wurde offensichtlich, dass diese Quoten in zahlreichen Studienfächern zu niedrig sind, also zu viele Personen ein Studium beginnen, dieses aber durch systemimmanente (fehlende Betreuung, schlechte Studienbedingungen, Finanzierungsprobleme etc.) oder persönliche Gründe (einseitige Information vor dem Studium, geringe Studieneignung für das gewählte Fach) abbrechen. Hohe Investitionen in die Ausbildung, ob privater oder volkswirtschaftlicher Natur, gehen so größtenteils verloren.

Im Mittelpunkt der Betrachtungen stand zudem der Eignungsbegriff (siehe Abschnitt 2.2.1 zur Personen-Umwelt-Passung). Es wurde beschrieben, dass eine Ausbildung immer dann mit einer hohen Wahrscheinlichkeit erfolgreich abgeschlossen werden kann, wenn das Stärken- und Schwächenprofil einer Person mit den Anforderungen eines Studienfaches zusammenpasst. Mit einem Studienabbruch oder einem Fachwechsel ist häufig – entgegen landläufiger Meinung – nicht eine generelle Nicht-Eignung für ein Studium verbunden, zeigen doch gerade Fachwechsler, dass sie für ein anderes Fach sehr wohl die passende allgemeine Eignung besitzen und dann das zuletzt gewählte Studium erfolgreich beenden (Spies, 1999). Ein zentrales Problem liegt somit

in vielen Fällen im Prozess der Studienfachwahl (siehe Abschnitt 2.1), eben im Abgleichen der fachspezifischen Eignung mit den Anforderungen begründet.

Gerade im Bereich der Informatik sind die Studienanfängerzahlen in der zweiten Hälfte der 1990er Jahre enorm angestiegen (Statistisches Bundesamt, 2001). Die Signale des Arbeitsmarktes, aber auch die markanten Stimmen der Politik, der Fachverbände oder Institute (Lückefett & Thomas, 2001; Petersen & Wehmeyer, 2001) zeigten eindrücklich ihre Wirkung (Heublein & Sommer, 2002). Seit 1975 bis 2000 hat sich die Anzahl der Studienanfänger fast verzwanzigfacht (Statistisches Bundesamt, 2006). Seit diesem Höchststand haben die Anfängerzahlen wieder kontinuierlich abgenommen, verbleiben aber auf einem hohen Niveau: 2006 begannen 28 357 Personen ein Studium in diesem Fach.

Nachweislich haben viele Studienanfänger dieser Boom-Jahre verstärkt nicht ihr besonderes Fachinteresse oder die fachspezifische Eignung, sondern die guten Arbeitsmarktchancen und eine oberflächliche Fachbeschreibung als ausschlaggebend für ihre Fachwahl angeführt (Heublein & Sommer, 2002). Offensichtlich basierte der rapide Anstieg der Studienanfängerzahlen auf einer Ausweitung der Zielgruppe und einer damit verbundenen Verlagerung der häufigsten Studienmotivationen, was sich wiederum in den Kennzahlen Studiendauer und Studienabbruchquote niedergeschlagen hat. Gleichzeitig haben die Informatik-Institute mit Überlastung kämpfen müssen, so dass ein gutes Betreuungsverhältnis zwischen Student und Dozent selten gegeben war. Die Studienzeiten in der Informatik sind seit 1996 kontinuierlich gestiegen (Wissenschaftsrat, 2001). So haben nur 4,2 % der Absolventen 1998 ihr Studium in der Regelstudienzeit von 9 Semestern beendet. Gleichsam sank die Erfolgsquote in den so genannten „Boom-Jahren“ nachweislich (Heublein, Schmelzer, Sommer & Spangenberg, 2002).

Diverse Möglichkeiten wurden in den vorigen Kapiteln beschrieben, wie eine Studienwahl optimiert beziehungsweise korrigiert und wie besonders die Erfolgsquoten durch eine stärkere Berücksichtigung der individuellen Passung erhöht werden könnten. Allen voran wurden zwei Herangehensweisen erwähnt, die es erlauben, möglichst viele Geeignete (siehe Abschnitt 2.2.1 und 2.2.2) zuzulassen und Ungeeigneten Alternativen aufzuzeigen: eine Unterstützung der *Selbstselektion* durch Studienberatung (siehe Abschnitt 2.4) und eine *Fremdselektion* durch die Hochschule (siehe Abschnitt 2.3).

Nach der Novelle des Hochschulrahmengesetzes (siehe Abschnitt 2.3) und der entsprechenden Landesgesetze sind Hochschulen gefordert, einen Teil der Studienplätze nach selbstbestimmten Kriterien zu vergeben – und zwar im Rahmen der so genannten Hochschulauswahlverfahren (HAV), auch Auswahlverfahren der Hochschulen (ADH) genannt (über die Ausgestaltung informieren Heine, Briedis, Didi, Haase & Trost, 2006). Somit kann zwar das spezifische Anforderungsprofil eines Fachbereichs definiert werden, auf der anderen Seite ist der Fachbereich dann meist selbst für die Qualitätskontrolle des Auswahlverfahrens zuständig. Gerade hinsichtlich der

großen Bedeutungen dieser Entscheidungen (daher wird auch häufig von *high-stakes*-Verfahren gesprochen) ist dies ein wichtiges und zugleich ein schwieriges Unterfangen.

Verfahren zur Fremdselektion sind nur dann rechtlich zulässig, wenn mehr Studienplatzbewerber als Studienplätze vorhanden sind (Bahro & Berlin, 2003). In den anderen, nicht reglementierten Studienfächern existieren aber teilweise eklatantere Zustände mit niedrigen Erfolgsquoten (Heublein, Schmelzer & Sommer, 2005). Eine Abhilfe gerade für diese Fächer könnte in der Unterstützung der individuellen Selbstselektion liegen. Diese bietet den Interessierten eine zusätzliche Beratungsleistung und das ohne den hohen Aufwand einer Fremdselektion.

Der Begriff der *Selbstselektion* (siehe Abschnitt 2.4) beschreibt zuallererst die unbeschränkte Studienfachwahl durch das Individuum, die in allen Studienfächern eine wichtige Rolle einnimmt. In den nicht beschränkten Studienfächern können die Interessenten – mehr oder weniger intuitiv und verzerrt – die Passung zwischen Studienfach und der eigenen Person abschätzen (siehe Abschnitt 2.1) und so die resultierende Wahl treffen. Somit existiert nach dem individuellen Entschluss kein Korrektiv. Neben den entsprechend hohen Abbruchquoten hat diese Orientierungslosigkeit noch eine weitere Auswirkung: Die dabei entstehenden Unsicherheiten bei den Schulabsolventen zeigen sich unter anderem auch in den im internationalen Vergleich geringen Übergangsquoten von der Schule zur Hochschule. Vielen Abiturienten, die sich trotz guter Noten gegen ein Hochschulstudium entscheiden, sind nicht unbedingt das deutsche Hochschulsystem oder gar die späteren Berufsfelder zu unattraktiv, vielmehr erscheint vielen die Entscheidung für eine akademische Ausbildung zu unsicher – und bei steigenden Studiengebühren aus ihrer gegenwärtigen finanziellen Lage auch noch zu teuer.

Die besondere Bedeutung der Selbstselektion und deren Möglichkeiten, aber auch Grenzen, stehen daher im weiteren Verlauf dieser Arbeit im Mittelpunkt. In Studie 1 (siehe Abschnitt 5) wird die letzte Version einer Testbatterie zur Studienberatung vorgestellt, die Entwicklungsgeschichte, die Testgüte und die Faktorenstruktur berichtet. Dabei wird insbesondere betrachtet, ob es gelungen ist, kurze und dennoch reliable Tests und Fragebogenskalen zu entwickeln.

In der Studie 2 (siehe Abschnitt 6) werden die Profilunterschiede zwischen Personen untersucht, die unterschiedliche Studienfächer studieren möchten. Dabei wird betrachtet, welche Tests am besten geeignet sind, die Unterschiede in der Neigung, aber auch in der Eignung aufzudecken. Bilden die Tests und Fragebogen nicht nur die allgemeine Studierfähigkeit ab, so sollten sich Unterschiede in den Testergebnissen von Studieninteressierten der Informatik/Elektrotechnik und der Geisteswissenschaften auf der anderen Seite zeigen. Der Abschnitt 6.2 beschreibt die Fragestellungen dieser Studie genauer.

In der Studie 3 (siehe Abschnitt 7) wird erörtert, inwieweit die Selbstselektion mit Verfahren verbessert werden kann, die ebenso wie die Fremdselektion auf psychometrischer Eignungs-

und Neigungsdiagnostik beruht. Die Kriteriumsvalidität, insbesondere die inkrementelle Validität über die Abiturdurchschnittsnote hinaus, steht somit im Mittelpunkt der Betrachtungen. Nicht nur Auswahlinstrumente, sondern auch Beratungssysteme müssen gültige Vorhersagen der Zukunft treffen können. Immerhin entscheiden beide Systeme über die Zukunft von Studieninteressierten, einmal direkt über die Fremdselektion und einmal indirekt über die Selbstselektion.

In Studie 4 (siehe Abschnitt 8) wird untersucht, wie *guidance at a distance* (siehe Abschnitt 2.5) in Deutschland von der Zielgruppe akzeptiert wird und welche subjektiven Erfolge in der Studienberatung erzielt wurden. Diese Studie verdeutlicht somit, inwieweit es dem System nicht nur gelungen ist, reliable und valide zu messen, sondern auch möglichst viele Studieninteressierte für eine Teilnahme zu gewinnen. Ein webbasiertes Studienberatungssystem kann natürlich erst dann Einfluss auf die mittlere Studienzufriedenheit und auf die Abbruchraten ausüben, wenn pro Studienjahrgang zahlreiche Studieninteressierte ihre Studienfachwahl durch eine Teilnahme gefestigt haben.

Kapitel 5

Studie 1: Entwicklung und Evaluation einer Testbatterie zur Studienberatung

5.1 Einleitung

An der RWTH Aachen stiegen – wie nahezu an allen Universitäten in Deutschland (Statistisches Bundesamt, 2001) – Ende der 1990er Jahre nicht nur die Bewerberzahlen in der Informatik, sondern ebenfalls die Abbrecherquote (Weber, 2003). Mit dem Ziel, langfristig die Studienerfolgsquoten wieder in den Fächern Informatik und Elektrotechnik zu steigern, wurde ein entsprechender Maßnahmenkatalog in die Zielvereinbarungen der RWTH Aachen mit dem Land Nordrhein-Westfalen aufgenommen (Dezernat Planung Entwicklung und Controlling der RWTH Aachen, 2002). Dabei wurde unter anderem festgeschrieben, die Studienberatung für die Informatik zu verbessern und die Selbstselektion der Studienplatzinteressenten mithilfe von frei zugänglichen und freiwillig durchführbaren Tests zu unterstützen (siehe Abschnitt 2.4 und 2.5). Zu häufig waren Informatik oder Ingenieurwissenschaften gewählt worden, weil die Absolventen auf dem Arbeitsmarkt gefragt waren (Heublein, Schmelzer, Sommer & Spangenberg, 2002), zu wenig war vielen Studieninteressierten über die Studieninhalte bekannt, zu wenig über ihre Eignung und Neigung für die Fächer.

Gemäß dieser Vereinbarungen wurde in zwei Diplomarbeiten (Weber, 2003; Zimmerhofer, 2003) und dieser vorliegenden Arbeit ein webbasiertes Studienberatungssystem insbesondere für den Studiengang Informatik entwickelt. Bezeichnet wurde dieses System mit dem Namen *Self-Assessment* (zur Definition des Begriffs siehe Abschnitt 2.5). Auf Basis einer umfassenden Anforderungsanalyse mit Dozenten, Studierenden und Studienberatern (siehe zum Ablauf und zu den Ergebnissen Zimmerhofer, 2003) wurden einzelne Prädiktoren aus den bereits vorgestellten (siehe Abschnitt 3.3 für die kognitiven und Abschnitt 3.4 für die nicht-kognitiven Bestandteile) ausgewählt und entsprechende Verfahren komplett neu konzipiert. Der Aufbau des Verfahrens wird in Abschnitt 5.3.2 genauer spezifiziert.

Nach einer Phase der Konzeption, Testentwicklung und technischen Umsetzung (siehe Abschnitt 5.3.3) durch den Autor dieser Arbeit können Interessierte seit dem 2. August 2002 auf die Internetseite <http://www.assess.rwth-aachen.de> zugreifen und das Beratungssystem unentgeltlich nutzen. Die Internettestung dauert in der letzten Testrevision inklusive der Instruktion etwa 90 Minuten. Die Teilnahme kann unabhängig vom Alter, vom Studieninteresse, vom Wohnort etc. erfolgen. Technischer Support wurde vom Autor direkt per E-Mail geleistet.

5.2 Fragestellung

In dieser ersten Studie steht die Vorstellung der Testbatterie mit dem jeweiligen Konstruktionshintergrund (siehe insbesondere Abschnitt 5.3) und der Messintention im Vordergrund. Aus Platzgründen kann hier allerdings die Konstruktion nicht vollständig seit Anbeginn der Konzeption – also seit dem Entwicklungsbeginn im Jahr 2002 – bis zur jetzigen vierten Revision dargestellt werden. Für eine umfangreichere Darstellung der ersten Revision, aber auch der durchgeführten Anforderungsanalyse und deren Ergebnisse kann auf die Arbeiten von Viola Weber (2003) und die des Autors (Zimmerhofer, 2003; Zimmerhofer, Heukamp & Hornke, 2006) verwiesen werden.

Neben der Vorstellung der Skalen werden hier insbesondere die Testgüte der vierten, also letzten Testrevision näher betrachtet sowie die internen Konsistenzen, die Interkorrelationen der Skalen und Faktorstruktur erläutert. Im Anhang (siehe Abschnitt 11) werden die Itemtexte samt Gütekriterien umfänglicher berichtet.

5.3 Methoden

5.3.1 Stichprobe

Die Stichprobe der Studie 1 umfasst alle Testungen, die zwischen August 2002 und April 2006 komplett abgeschlossen wurden. In diesen Jahren sind insgesamt Daten bei 11 694 Testdurchgängen entstanden. Diese Zahl ist bereits um Datensätze mit Übertragungsfehlern beziehungsweise um Dubletten¹ bereinigt. Da die weitere Analyse auf der Basis von Ersttestungen vorgenommen werden soll, wurden Retestungen ebenfalls ausgeschlossen. Dazu wurden zwei unterschiedliche Kriterien verwandt: Erstens wurden Teilnehmer danach gefragt, ob der jeweilige Testdurchlauf eine Retestung darstellt. Zweitens mussten sich alle Teilnehmer registrieren,

¹ Mit Dubletten sind doppelte Einträge derselben Testdurchführung in der Datenbank gemeint. Entstanden sein können diese durch ein mehrfaches Absenden der Testdaten an die Datenbank durch schnelle Klicks auf den Absende-Button.

so dass Retestungen unter demselben Login automatisch eliminiert werden konnten. Aufgrund dieser beiden Vorgehensweisen reduziert sich der Datensatz auf 10 708 Testungen.

Da bei Testungen, die frei im Internet stattfinden, davon ausgegangen werden muss, dass diese zu einem gewissen unbekanntem Prozentsatz nicht ernsthaft durchgeführt werden, sind diverse Anstrengungen zur Detektion von solchen „aberranten“ Antworten unternommen worden. Anders als bei kontrollierten Testungen in Gruppen oder gar in Einzelsituationen besteht bei den so genannten *self-administered questionnaires* im Internet das Problem, dass die Probanden völlig unbeobachtet Aufgaben bearbeiten. Dabei bleiben sowohl die Bedingungen der Testteilnahme (ablenkende Faktoren, z. B. Alkoholgenuß, besondere Lärmquellen etc.) als auch die Ernsthaftigkeit unklar, mit der gearbeitet wurde. Möglicherweise vermindert eine Anonymisierung bei webbasierten Testungen zwar den Bias der sozialen Erwünschtheit (Joinson, 1999; Richman, Kiesler, Weisband & Drasgow, 1999; Rosenfeld, Booth-Kewley, Edwards & Thomas, 1996), die empfundene Verantwortlichkeit für die erfolgte Beantwortung nimmt aber ebenfalls ab (Krosnick, 1991). Entsprechend wird hier davon ausgegangen, dass ein schwer bestimmbarer Anteil von Personen unter den beschriebenen Bedingungen einer anonymen, freiwilligen Internettestung Zeit und kognitive Ressourcen bei der Bearbeitung einsparen (sich also einfach weniger Gedanken über das Aufgabenmaterial machen), dennoch aber das Verfahren und die Rückmeldung explorieren möchten. Nach Krosnick (1991) lässt sich dabei ein Kontinuum beschreiben, dass von einer eingeschränkten Vereinfachung des kognitiven Prozesses (*weak satisficing*) bis zur nahezu kompletten Reduktion der Anstrengung (*strong satisficing*) geht. Entsprechend bearbeiten manche Personen gerade kognitiv aufwändige Subtests nur wenig konzentriert oder – im Extremfall – komplett ohne kognitiven Aufwand². Mit diesem Vorgehen reduzieren sie zwar ihren Aufwand während der Bearbeitung, können aber dennoch den Testlauf beenden, um vielleicht ein interessantes Feedback etc. zu erhalten oder an einem Gewinnspiel teilzunehmen.

In der Planungsphase dieser Arbeit wurden mehrere Möglichkeiten zur Detektion der nicht validen Datensätze eruiert, sowohl auf Basis der klassischen und der probabilistischen Testtheorie als auch auf Basis von Mehrebenenanalysen (siehe zur Übersicht Zimmerhofer & Hornke, 2005a). Als gangbares Kriterium wird hier die Bearbeitungszeit herangezogen, welche sich als überaus praktikabel gezeigt hat. Um nicht Datensätze von besonders leistungsstarken und daher schnell antwortenden Personen zu eliminieren, wird pro Itemtyp eine konservative Zeitgrenze festgelegt, die von der in Pretests benötigten Lesedauer abhängt. Maximal werden in den hier berichteten Datensätzen 10 % der Testfälle separiert. Ähnliche Quoten werden erzielt, wenn Datensätze eliminiert werden, deren Latenzzeiten auf Subtestebene den Median minus einer Standardabweichung unterschreiten. Nach dem Data-Screening reduziert sich die Stichprobengröße auf $n = 9\,227$. Datensätze, die bei der Kriteriumsvalidierung entstanden (siehe Abschnitt 7.4) sind, werden für die Studie 1 nicht verwandt. Entsprechend stehen schlussendlich für alle

² In der Marktforschung wird ein vergleichbares Antwortverhalten „non-attitude-“ oder „no-opinion-response“ genannt (Krosnick et al., 2002).

eingesetzten Verfahren außer der Handlungskontrolle (*HK*) 8 633 Testdaten zur Verfügung. Für *HK* steht aufgrund einer später erfolgten, letzten Testrevision nur eine Stichprobe von 6 181 Personen zur Verfügung.

Im Mittel sind die letztlich hier betrachteten Teilnehmer 20,9 Jahre alt ($SD = 3,98$). 24,8 % sind weiblich, 75,2 % männlich. 96,9 % kommen aus Deutschland, 46,1 % aus Nordrhein-Westfalen. Für 9,0 % der Teilnehmer ist Deutsch nicht die Muttersprache.

Um die Zugriffszahlen auf die Webseite durch die Zielgruppe zu erhöhen, wurden mehrere Maßnahmen verfolgt: Auf der einen Seite wurden Schüler direkt für die Beratungsleistung des Self-Assessments und dessen Bedeutung für die Studienfachwahl sensibilisiert. Dies geschah an Tagen wie dem *Dies Academicus* beziehungsweise dem *Tag der Informatik* der RWTH Aachen in den Jahren 2003 bis 2006. Zudem wurden Besuche an Schulen im Rahmen des Projektes *Science Truck* der zentralen Studienberatung (siehe <http://www.rwth-aachen.de/sciencetruck>) durchgeführt. Auch wurde Informationsmaterial über das Beratungssystem postalisch an alle Gymnasien des Aachener Kreises gesendet. Bei diesen Anstrengungen standen Lehrer als Multiplikatoren im Mittelpunkt, konnte man doch über sie vergleichsweise leicht zahlreiche Schüler von den Vorteilen des Systems überzeugen. Ferner wurden die Philologenverbände, Journalisten aus den Bildungs-, Wissenschafts- und IT-Ressorts und auch Webmaster großer deutscher Internetportale kontaktiert und um eine Erwähnung in Printmedien oder Webseiten gebeten. Mit diesen Anstrengungen wurde eine Veröffentlichung in diversen Printmedien und auf Internetseiten erreicht (zum Beispiel <http://www.vdi.de>, <http://www.arbeitsagentur.de>, <http://www.zeit.de>, <http://www.uni-essen.de/isa>, <http://www.heise.de>). Einträge in zahlreichen Suchmaschinen – bei Spidersuchmaschinen, wie zum Beispiel <http://www.google.de> und auch bei Katalogsuchmaschinen, wie zum Beispiel <http://www.dmoz.org> – wurden ebenfalls erstellt.

Für eine Teilnahme geworben wurde primär mit der Beratungsleistung, also mit einem ausführlichen Feedback zur Studienorientierung und zu den eigenen Stärken und Schwächen hinsichtlich einer Studienaufnahme. Ferner wurde darauf hingewiesen, dass mit einer Teilnahme auch Erfahrung mit eignungsdiagnostischen Instrumenten gesammelt wird. Zudem wurden kleinere Geschenke (Amazon-Büchergutscheine, SUSE-Linux-Pakete, Bücher etc.) als weiterer Anreiz verlost. Die Marketingbemühungen zielten darauf ab, möglichst viele Teilnehmer der Zielgruppe zu gewinnen, aber auch gleichzeitig die Dropout-Rate zu minimieren. So wurden die kleineren Gewinne nur unter den Personen verlost, die einen Testdurchlauf ernsthaften bearbeitet und abgeschlossen haben. Die Gewinnwahrscheinlichkeit konnte noch erhöht werden, indem die Personen den Akzeptanzfragebogen (siehe Abschnitt 8) ausfüllten. Beide Anreizformen werden von Reips (2002) empfohlen.

5.3.2 Verwandte Tests

Da das Self-Assessment als ein Element einer freiwilligen Studienberatung verstanden wird und entsprechende Instrumente komplett frei zugänglich im Internet sein sollten, konnten aus Lizenzgründen keine kommerziellen Fragebogen und Tests verwandt werden. Alle Verfahren sind daher vom Autor in den Jahren 2002 bis 2006 entwickelt worden, im Jahr 2002 in Zusammenarbeit mit Viola Weber (2003). Die Auswahl der einzelnen Prädiktoren des Studienerfolgs wurden mittels theoretischer Überlegungen zur Validität (siehe Abschnitt 3.3 für die kognitiven und Abschnitt 3.4 für die nicht-kognitiven Bestandteile) und einer umfassenden Anforderungsanalyse (beschrieben in Zimmerhofer, 2003) getroffen.

Berufliches Interesse (*INT*)

Wie bereits beschrieben (siehe Abschnitt 3.4.3) spielt das berufliche Interesse eine entscheidende Rolle bei der Studien- und Berufswahl Heranwachsender (Heine & Willich, 2006; Lewin, Heublein, Schreiber, Spangenberg & Sommer, 2002). Auf dieser Grundlage scheint der Einsatz eines entsprechenden Interessenfragebogens gerade für die Studienfachwahl entscheidend, soll doch dadurch nicht nur das Interesse bezüglich des Studienfaches sondern gerade auch bezüglich des späteren Berufs aufgedeckt werden. Der hier im Self-Assessment eingesetzte Fragebogen basiert auf der bereits beschriebenen Theorie nach Holland (1997) (siehe Abschnitt 3.4.3) und wurde vom Autor dieser Arbeit im Rahmen der Diplomarbeit (Zimmerhofer, 2003) in einer ersten Fassung entwickelt und nachher optimiert, um einen reliablen Fragebogen bei kürzerer Testdauer zu erhalten. Dazu wurden sowohl Itemanalysen nach klassischer als auch nach probabilistischer Testtheorie verwandt. Die sechs Dimensionen der Theorie werden in der hier betrachteten vierten Testrevision jeweils durch 10 Items abgebildet, die berufs- beziehungsweise studienbezogene Tätigkeiten beschreiben.

Bereits bei der Entwicklung der Skalen, aber auch bei der weiteren Testoptimierung wurde die besondere Zielgruppe des Verfahrens berücksichtigt: Mehr als andere Interessenfragebogen (zum Beispiel Bergmann & Eder, 1992 oder Jörin, Stoll, Bergmann, Eder & Bundesagentur für Arbeit, 2004) ist dieses Verfahren darauf abgestimmt, dass die Teilnehmer die allgemeine Hochschulreife anstreben beziehungsweise schon erlangt haben. Im Mittelpunkt der Bemühungen stand daher, Tätigkeiten als Operationalisierungen der sechs Dimensionen zu finden, die dem Erfahrungsraum der Studieninteressierten entsprechen und akademische Berufe beschreiben. Berufsbilder und Tätigkeiten, die primär einer dualen Ausbildung zugeordnet werden können, wurden bei der Fragebogenkonstruktion nicht berücksichtigt. Aufgrund der klar umrissenen Zielgruppe des Verfahrens ist indirekt auch die Altersstruktur der Teilnehmer festgelegt, die bei etwa 16 Jahren beginnt und bei 24 Jahren endet.

In Vorbereitung der ersten Testversion wurde – unter anderem mit Schülern und Erstsemesterstudenten – die leichte Verständlichkeit der Items überprüft. Vergleichbar mit anderen aktuellen Interessenfragebogen (zum Beispiel dem AIST-R von Bergmann und Eder, Bergmann & Eder, 2005) wurde bei der Konstruktion ein mittleres Abstraktionsniveau der Tätigkeitsbeschreibungen (Todt, 1978) anvisiert. Damit sollte auf der einen Seite eine Generalisierung auf nicht explizit erfragte Tätigkeitsbereiche ermöglicht werden. Auf der anderen Seite sollten die Formulierungen aber noch gut vorstellbar und somit konkret genug sein.

Die Einschätzung der Tätigkeiten wurde anhand einer fünfstufigen Rating-Skala (von „total langweilig“, über „nicht so interessant“, „interessant“, „sehr interessant“ bis „super interessant“) vorgenommen. Dabei mussten 90 Tätigkeiten in der ersten Testrevision beziehungsweise später 60 Tätigkeiten bewertet werden. Fehlende Werte (sogenannte *missings*) konnten konstruktionsbedingt aus technischen Gründen im Datensatz nicht auftreten, da die Beantwortung jedes Items obligatorisch war. Die Bearbeitung war nicht zeitbegrenzt. Die Testdauer betrug inklusive der Instruktionszeit im Mittel etwa sechs Minuten.

Jeder Teilnehmer hat nach Abschluss des Self-Assessments eine Rückmeldung erhalten, die für den Bereich der beruflichen Interessen folgende Bestandteile enthält:

1. eine kurze und leicht verständliche Einführung in die Theorie Hollands,
2. die drei präferierten Holland-Dimensionen,
3. mögliche, dazu passende Studienfächer,
4. einen normativen Vergleich mit einer Stichprobe ($n = 759$), die sich aus Interessierten der Fächer Informatik, Elektrotechnik und Technische Informatik zusammensetzt,
5. der Personen-Umwelt-Passung auf Basis des *C-Index* nach Brown und Gore (1994),
6. eine grafische Visualisierung der Personen-Umwelt-Passung in Form eines Spinnennetzes als Hexagon.

Zur weiteren empirischen Überprüfung werden nicht nur die Einzelskalen des Interessenfragebogens betrachtet, sondern auch ein Index, der die Passung der Person zur Umwelt eines IT-Studiengangs quantifiziert. Verwandt wird dabei der C-Index nach Brown und Gore (1994). Zur Festlegung der Umwelt wurde zu Beginn der Konstruktion auf ein Expertenrating von Bergmann und Eder (1992) zurückgegriffen (siehe zur weiteren Erläuterung Abschnitt 3.4.3). Diese Festlegung der Umwelt auf *I-R-E* deckt sich mit durchgeführten empirischen Überprüfungen zur ersten Testrevision (Zimmerhofer, 2003).

Extrinsische Lernmotivation (ELM)

Wie bereits im Abschnitt 3.4.4 beschrieben kann aufgrund empirischer Ergebnisse davon ausgegangen werden, dass nicht nur alleine die intrinsische (hier als Interesse operationalisiert, siehe

Abschnitt 3.4.3), sondern auch die extrinsische Lernmotivation eine wesentliche Motivationsquelle im Studienalltag darstellt. Bei der Umsetzung des theoretischen Konzepts in der ersten Testrevision wurden sowohl Items zur Orientierung an der Selbstbewertung, der Fremdbewertung und an Oberzielen verwandt (Heckhausen, 1989). Entsprechend wird angenommen, dass Personen dadurch motiviert werden, den eigenen Ansprüchen, also der individuellen Bezugsnorm folgend, zu genügen und über die Zeit bessere Leistungen zu erbringen (*Selbstbewertung*). Natürlich motiviert darüber hinaus auch Lob, ob ausgesprochen durch Gleichaltrige oder auch durch Lehrer, Dozenten beziehungsweise Eltern als *Fremdbewertungen* (Schiefele & Urhahne, 2000). Eine weitere wichtige Quelle sind anvisierte *Oberziele*, also zum Beispiel das Erreichen eines besonderen Berufs. Da diese drei Motivationsquellen unabhängig von den Inhalten der ausgeführten Tätigkeit aktivieren, sind sie gerade im Studium aufgrund der hohen Eigenverantwortlichkeit und der geringen Kontrollmechanismen essentiell. Sie können nämlich immer auch dann antreiben, wenn ein bestimmter Stoffinhalt nicht intrinsisch motiviert.

Genauso wie der Fragebogen zum beruflichen Interesse ist dieser Fragebogen zur *ELM* in der ersten Revision im Rahmen der Diplomarbeit des Autors (Zimmerhofer, 2003) entwickelt worden. Dabei wurden alle drei oben genannten Zielformen gleich häufig im Fragebogen abgebildet. Nach ersten Datenerhebungen wurde der Fragebogen auf Grund probabilistischer und klassischer Gütekennwerte um ein Drittel verkürzt, um – gerade für das Internet essentiell – eine effizientere Testung zu ermöglichen. Die letzte, hier näher betrachtete Testrevision umfasste zehn Items mit einer vierstufigen Ratingskala („lehne vollständig ab“, „lehne eher ab“, „stimme mit Einschränkung zu“, „stimme voll zu“). Die Gefahr, dass aus einer sozialen Erwünschtheit heraus zaghafte (also mittlere) Antworten wie „weiß-nicht“ oder „teils-teils“ gegeben werden, war somit gemindert. Jedes Item musste – ebenso wie bei allen anderen Items im Self-Assessment – beantwortet werden. Die Testdauer betrug inklusive Instruktion im Median etwa drei Minuten. Die Bearbeitungsdauer war nicht beschränkt.

Die Ergebnisrückmeldung bestand aus

1. einer Einführung zum Konstrukt der extrinsischen Lernmotivation,
2. einer Verbalisierung des individuellen Rohwertes,
3. einem normativen Vergleich mit einer Stichprobe ($n = 759$), die sich aus Interessierten der Fächer Informatik, Elektrotechnik und Technische Informatik zusammensetzt inklusive einer Angabe des Prozentrangwertes und
4. einer grafischen Visualisierung des Ergebnisses.

Studienbezogene Selbstwirksamkeit (SW)

Im Kapitel 3.4.5 wird die Bedeutung der eigenen Selbstwirksamkeitserwartungen für die Prognose des Studienerfolgs diskutiert. Dabei steht im Mittelpunkt, dass Personen bestimmte

Verhaltensweisen mit einer höheren Wahrscheinlichkeit zeigen, wenn sie sich in der Lage sehen, dieses auch erfolgreich abschließen und den anvisierten Nutzen daraus ziehen zu können. Lernverhalten wird also dann gezeigt, wenn die betroffene Person davon ausgeht, das Lernpensum auch zu erreichen und – genauso wichtig – das Gelernte auch in der Prüfung etc. mit Erfolg anwenden zu können. Verwandt für die Messung dieses Konstruktes wird der im Rahmen des Projektes vom Autor entwickelte Fragebogen (zur Entwicklung der ersten Testrevision siehe Zimmerhofer, 2003). Die Items sind mit einer Ratingskala versehen, die der des Fragebogens zur extrinsischen Lernmotivation gleicht. Nach einer umfassenden Itemanalyse wurde die 15 Items umfassende erste Testversion auf eine kürzere 10-Item-Version gekürzt, die hier näher untersucht wird. Die Formulierung wurde gemäß den Empfehlungen von Schwarzer (1999) für Fragebogen der Selbstwirksamkeit vorgenommen. Die Bearbeitungszeit betrug durchschnittlich drei Minuten. Beschränkt wurde sie nicht. Jedes Item musste bearbeitet werden, um weiter an der Testbatterie teilnehmen zu können.

In dem hier näher beschriebenen Fragebogen beziehen sich alle Itemformulierungen auf Lernverhalten. Entsprechend wird thematisiert, in wie weit sich Personen im akademischen Umfeld (insbesondere beim Vorbereiten auf Prüfungen) als selbstwirksam einschätzen. Diese Einschränkungen auf einen spezifischen Gegenstandsbereich stehen im Einklang mit der Annahme, dass Selbstwirksamkeitserwartungen von Individuen selten genereller sondern eher situationsspezifischer Natur sind (Scholz, Dona, Sud & Schwarzer, 2002). Dabei hängt das optimale Niveau der Generalität davon ab, was und in welchem Abstraktionsgrad auf der Seite des Kriteriums prognostiziert werden soll (Bandura, 1997). Diese eher allgemein gültige Empfehlung entspricht dem Ansatz des Brunswikschen Linsenmodells (siehe dazu auch Abschnitt 3.3.2 oder Wittmann, 2002). In dem hier beschriebenen Fall soll weder alleine das Verhalten in einer spezifischen Situation (Prüfungsvorbereitung in einem bestimmten Fach) vorhergesagt werden, noch ist es Ziel, eine allgemeine Prädiktion von Verhalten außerhalb des universitären Kontextes zu ermöglichen. Die Ergebnisrückmeldung folgt dem Aufbau der extrinsischen Lernmotivation.

Studienbezogene Handlungs- beziehungsweise Lageorientierung (HK)

Entsprechend der Theorie der Handlungskontrolle (Kuhl, 1983; Kuhl, 1987) und der Umsetzung im HAKEMP 90 von Kuhl (o.J.) bilden drei Dimensionen das Konstrukt: Handlungsorientierung nach Misserfolg (*HOM*), Handlungsorientierung bei der Handlungsplanung (*HOP*) und Handlungsorientierung bei der Tätigkeitsausführung (*HOT*). Die Bedeutung dieser drei Dimensionen für das studentische Umfeld – insbesondere für die Prüfungsvorbereitung – ist eindrücklich (siehe Abschnitt 3.4.6 zu empirischen Ergebnissen): Sowohl während der Planung von Prüfungsvorbereitungen als auch bei der Prüfungsvorbereitung selbst sowie nach missglückten Prüfungen scheint es wichtig, die zu erreichenden Ziele nicht aus den Augen zu verlieren

(Aspekt der *Volition* beziehungsweise des *Willens* zur Zielerreichung). Eine Starre, die so genannte Lageorientierung, sollte ebenfalls vermieden werden.

Analog zu den bereits beschriebenen Skalen ist auch bei der Konstruktion der Items zur Handlungs- und Lageorientierung auf spezifische Formulierungen bezüglich des akademischen Lernumfeldes geachtet worden. So werden die Aspekte der Handlungs- beziehungsweise der Lageorientierung nicht generell, sondern insbesondere während der Planung und Initiierung von Lernverhalten, während des Lernverhaltens selbst und nach einem Misserfolg in diesem Umfeld erfragt.

In der Ursprungsversion des Fragebogens, die wiederum im Rahmen der Diplomarbeit des Autors entstanden ist, wurden bei jedem Item zwei ausformulierte Antwortmöglichkeiten verwendet, die der Handlungsorientierung beziehungsweise der Lageorientierung zugeordnet waren. Die Entscheidung zugunsten eines dieser beiden Extreme war allerdings erst nach zwei Sekunden Bedenkzeit möglich. Diese erzwungene Latenzzeit verminderte die Gefahr, dass Personen eine Alternative unbeabsichtigt – zum Beispiel durch einen Doppelklick auf die vorhergehende Antwort – auswählen. Seit der zweiten Testrevision ist die zweistufige Skala auf eine insgesamt vierstufige erweitert worden. Dabei wurden die Extrempositionen (also die Ausformulierungen der Lage- beziehungsweise der Handlungsorientierung) der ursprünglichen Ratingskalen übernommen, die beiden mittleren Antwortkategorien entsprechend den Example Anchored Scales mit Abstufungen gekennzeichnet (Bortz & Döring, 1995). Mehrstufige Ratingskalen bei einem Fragebogen zur Handlungsorientierung haben U. Schiefele, Moschner und Hustegge (2002) im Rahmen der SMILE-Studie erfolgreich eingesetzt und dabei gute Messgenauigkeiten erzielt.

Die Ergebnismeldung für die Handlungskontrolle entspricht dem Vorgehen bei den bereits beschriebenen nicht-kognitiven Fragebogen.

Logisches Schlussfolgern (LOG)

Logisches Schlussfolgern wurde in den Anforderungsanalysen vor der Entwicklung der ersten Version des Self-Assessments für das betrachtete Fach Informatik als sehr wichtig erachtet (Zimmerhofer, 2003). Diese Einschätzung der interviewten Studenten und Dozenten deckt sich sowohl mit dem wissenschaftlichen als auch mit dem Alltagsverständnis, dass dieser Fähigkeitsbereich häufig als prototypisch für den Begriff Intelligenz angesehen wird (Rips & Conrad, 1983).

Das im Self-Assessment eingesetzte Verfahren beschäftigt sich insbesondere mit einer Form des logischen Denkens, dem deduktiven Schluss. Häcker und Stapf (1998) definieren diesen als „wahrheitsdefinite, nach den Regeln der formalen Logik mechanisch beweisbare Ableitung von Sätzen aus anderen, gegebenen Sätzen“ (S. 168). Dabei entstehen durch den deduktiven Schluss

keine völlig neuen Informationen. Der semantische Gehalt der Prämissen bleibt erhalten. Deduktives Denken ist wichtig beim Entscheiden über Ursachen von Ereignissen und Handlungen, dem Verfolgen von Argumenten, der Bewertung von Annahmen, Hypothesen und Fakten, dem Lösen von Problemen und der Entwicklung und Verfolgung von Plänen und Absichten (Weber, 2003). Entsprechend hat es eine entscheidende Bedeutung für die Entwicklung von Wissenschaft und Technik (Johnson-Laird & Byrne, 1991).

Das logische Schlussfolgern ist in zahlreichen Faktorenanalysen und in resultierenden Intelligenzstrukturmodellen als Faktor *Reasoning* enthalten, dass somit sowohl den deduktiven als auch den induktiven Schluss umfasst (Süß, 2003). Diese Aufnahme der Deduktion und Induktion als einen gemeinsamen Intelligenzfaktor – wie zum Beispiel in Thurstones Primärfähigkeiten – deckt sich mit dem allgemeinen Problem, dass faktorenanalytisch eine Trennung der beiden Aspekte nur in seltenen Fällen gelingt (Weber, 2003). Carroll (1993) nennt diesen Faktor *Sequential Reasoning*, dessen Bedeutung er insbesondere sieht „in tasks or tests that require subjects to start from stated premises, rules, or conditions and engage in one or more steps of reasoning to reach a conclusion that properly and logically follows from the given premises“ (S. 245).

Das ab der zweiten Testrevision verwandte Itemmaterial umfasste auf der Basis von 11 Items verschiedene Bereiche der Aussagen- und Prädikatenlogik. Mit dem Ziel, im Rahmen des Self-Assessments den Bezug zu den Studiengängen Informatik, Technische Informatik und Elektrotechnik zu betonen und somit auch eine höhere Akzeptanz zu erzielen, wurden einige Items darüber hinaus mit deutlichem Technik-Bezug formuliert.

Weitere Elemente der Itemkonstruktion waren (siehe Weber, 2003, S. 50):

- Verwendung von Negationen, die im Vergleich zu Aufgaben ohne Negationen als schwieriger gelten
- Berücksichtigung der verschiedenen bekannten Schlussformen (zum Beispiel Modus tollens) und ihrer typischen Fehler (wesentlich für die Entwicklung der Distraktoren)
- Bei relationalen Items:
 - Homogenität der Beschreibung (zum Beispiel ist „A geschieht vor B und B geschieht vor C“ leichter zu lösen als „A geschieht vor B und C geschieht nach B“, obwohl exakt die gleiche zeitliche Reihenfolge beschrieben wird)
 - Verwendung horizontaler, vertikaler und zweidimensionaler räumlicher Anordnungen (vertikale Problemstellungen werden meist als leichter eingestuft)

Das erste Item des Subtests wurde nicht ausgewertet, da dieses lediglich zur Eingewöhnung der Testteilnehmer diente. Die Bearbeitungszeit war auf 360 Sekunden pro Item beschränkt. Nach 300 Sekunden wurde eine Meldung eingeblendet, dass noch eine Minute an Bearbeitungszeit verblieb. Zudem wurden die Antwortmöglichkeiten erst nach 2 Sekunden freigeschaltet, um

versehentliches Antworten, aber auch unmotiviertes Weiterklicken möglichst zu vermeiden. Die Entwicklung der ersten Testversion wurde von Weber (2003) vorgenommen, die Weiteren bis zur heutigen vierten vom Autor dieser Arbeit.

Die Ergebnisrückmeldung für den Logiktest entspricht dem Vorgehen bei den bereits beschriebenen nicht-kognitiven Fragebogen. Um den Mehrwert für die Nutzer noch zu vergrößern, wurden die Lösungswege von drei falsch beantworteten Aufgaben erläutert.

Matrizentest (MAT)

Neben dem vorgenannten Verfahren zum deduktiven Denken wird in der Testbatterie des Self-Assessments noch ein Verfahren zur Erfassung des induktiven Denkens verwandt, das zudem als Maß für die allgemeine Intelligenz gilt (Jensen, 1998). Beim induktiven Denken steht der Schluss im Mittelpunkt, der „von beobachteten Gegebenheiten auf die (meist potentiell unendliche) Menge nicht beobachtbarer gleichartiger Gegebenheiten, also vom Besonderen auf Allgemeines oder von Beobachtungen auf Gesetzmäßigkeiten“ (Glaser, 1998, S. 394) vollzogen wird. Somit lässt sich die Induktion als wahrheitsvermehrender Erweiterungsschluss verstehen. Im Gegensatz zum Deduktionsschluss lässt sich der Induktionsschluss nicht durch eine einheitliche Sprache darstellen, durch mechanische Anwendung von Regeln beweisen oder begründen.

Bei dem hier verwandten Itemtyp besteht die Aufgabe für die Teilnehmer darin, eine Anordnung geometrischer Elemente in Form einer Matrix um ein fehlendes Element aus mehreren möglichen Antwortmöglichkeiten zu ergänzen. Dabei stellen die vorgegebenen spalten- und zeilenweisen Anordnungen der Elemente spezielle Fälle dar, in denen grundlegende Konstruktionsregeln (zum Beispiel Addition, Subtraktion, Identität, vgl. Hornke, Küppers & Etzel, 2000) erkannt werden müssen, damit die richtige Antwortoption gewählt werden kann.

Die Grundidee zu einem solchen Itemmaterial – den so genannten Matrizenitems – stammte von John Carlyle Raven, der unter Mithilfe von Lionel S. Penrose diese Form von Leistungstests vor etwa 70 Jahren entwickelte (Burke, 1958). Nach Raven, einem Schüler Spearman, sollte sein Verfahren der Messung der *innate eductive ability* (Raven, 1938, zitiert nach Burke, 1958) dienen, also der angeborenen Fähigkeit, in Komplexem Gesetzmäßigkeiten zu finden. Raven (1989) beschreibt dies ferner mit „the ability to educe correlates, the ability to generate high level schemata, which make it easy to handle complex events“ (S. 1). Davon getrennt wird ebenfalls eine Erfassung der *reproductive ability* angestrebt, „the ability to recall acquired information“ (Raven, 1989, S. 1). Spearman, der an der Entwicklung des ersten Matrizentests seines Schülers Raven beteiligt gewesen sein soll (Williams, Zimmerman, Zumbo & Ross, 2004), bewertet Matrizentests als beste Möglichkeit, nonverbal die allgemeine Intelligenz seines Zwei-

Faktoren-Modells zu messen (Spearman, 1946, zitiert nach Burke, 1958). Passend dazu klassifizieren Vernon und Parry entsprechende Verfahren als „an almost pure g test“ (Burke, 1958, S. 202; Mackintosh & Bennett, 2005).

In Cattells Erweiterung des Intelligenzmodells nach Spearman werden Matrizen tests der fluiden Intelligenz g_f ³ (Cattell, 1987), also den Basisprozessen des Denkens zugeordnet. Im Mittelpunkt steht, dass – zumindest der Theorie nach – die Leistungen nicht abhängig sind von Aspekten der Bildung und der Kultur. Messungen von g_f sollen primär mit Itemmaterial erfolgen, das entweder allen Teilnehmern bekannt (überlernt) ist oder für alle neues Material darstellt. Marshalek, Lohman und Snow (1983) bezeichnen Matrizen tests – unter anderem aufgrund des artifiziellen Itemmaterials – als prototypische Aufgaben für die fluide Intelligenz. Carroll (1993) konnte diese Ansicht in großen Faktorenanalysen belegen, die schlussendlich zum Three-Stratum-Modell führten. Dabei wiesen Matrizenaufgaben erwartungsgemäß substantielle Ladungen mit dem generellen Intelligenzfaktor g auf, dem g_f Faktor, dem induktiven Denken (*induction*) sowie dem *Sequential Reasoning* auf der untersten Ebene.

Im Self-Assessment wurden in der ersten Testrevision 15, von der zweiten bis zur dritten Testrevision 11 Matrizenitems mit jeweils acht Antwortoptionen verwandt, die entsprechend den Regeln von Hornke, Küppers und Etzel (2000) entwickelt wurden. Mit der letzten Testrevision wurden zwei neue Matrizenitems im oberen Schwierigkeitsbereich eingeführt, die hier im Weiteren aber nicht näher betrachtet werden. Die Entwicklung der ersten Testversion wurde von Weber (2003) vorgenommen, die Weiteren vom Autor dieser Arbeit. Vergleichbar mit dem Vorgehen bei anderen Leistungstests im Self-Assessment wurde auch beim Matrizen test das erste Item nicht ausgewertet, da dieses lediglich zur Eingewöhnung des Testteilnehmers dient. Die Bearbeitungszeit wurde auf 360 Sekunden pro Item beschränkt. Nach 300 Sekunden wurde eine Meldung eingeblendet, dass noch eine Minute verbleibt. Zudem wurden die Antwortmöglichkeiten erst nach 2 Sekunden freigeschaltet, um versehentliches Antworten möglichst zu vermeiden.

Die Ergebnisrückmeldung für den Matrizen test entspricht dem Vorgehen beim Logiktest.

Mathematische Fähigkeiten (MATH)

Gerade für den Bereich der Studiengänge der Informatik und der Ingenieurwissenschaften zeigen bisherige Anforderungsanalysen die Bedeutung mathematischer Aufgabenstellungen zur Vorhersage des Studienerfolgs (Blum, 1980; Zimmerhofer, 2003). Dabei sind weniger die Vorkenntnisse und bestimmte Fertigkeiten als die grundlegenden mathematischen Fähigkeiten relevant. Anders als bei mathematischem Wissen (zum Beispiel über bestimmte Formeln) wird

³ Cattell selbst beschreibt die fluide Intelligenz mit „purely general ability to discriminate and perceive relations between any fundaments, new or old“ (Cattell, 1943, S. 178).

beim Fähigkeitsbegriff von einem Trait gesprochen, der über eine längere Zeit Stabilität aufweist (Fleishman, Costanza & Marshall-Mies, 1999). Vorkenntnisse beschleunigen zwar den Neuerwerb von mathematischem Wissen, lassen sich andererseits aber auch zeitnah aufarbeiten, wenn die grundlegenden Fähigkeiten dazu vorhanden sind.

Was genau diese mathematischen Fähigkeiten umfassen und wie diese zu anderen verwandten Konstrukten abzugrenzen sind, ist wissenschaftlich umstritten, so dass Definitionen, aber auch die entsprechenden Tests als Operationalisierungen von Autor zu Autor beträchtlich differieren. Zudem scheinen – entsprechend der gesellschaftlichen Entwicklungen – auch Verschiebungen in den Definitionen stattzufinden: So wurden einfache zahlengebundene Probleme zunehmend durch komplexere Zahlen-(Relations-)Leistungen abgelöst, die ebenfalls eine numerische Basisfertigkeit erfordern (Rettig, Hornke & Schiff, 1989). Unabhängig von dieser Unklarheit wird eine Messung mathematischer Fähigkeiten in zahlreichen publizierten und unpublizierten Tests angestrebt: Heutige Verfahren werden meist durch Additions-, Subtraktions-, Multiplikations- oder Divisionsaufgaben, Textaufgaben, Umformungen, das Fortsetzen von Zahlenreihen, Setzen von Rechenzeichen oder das Lösen geometrischer Probleme gebildet (Weber, 2003).

Carroll (1993) charakterisiert das Konzept der mathematischen Fähigkeiten entsprechend dieser Einschränkungen als ein unexaktes, unanalysiertes, aber durchaus populäres Konstrukt, das ohne wissenschaftliche Bedeutung ist. Zudem kommt er zu dem Schluss, dass bei einer Lösung mathematischer Aufgaben immer mehrere Fähigkeiten zugrunde liegen. Faktorenanalytisch lassen sich dabei zwei unterschiedliche, gut belegbare Faktoren finden: *Numerical Facility* (auch kurz *Number* genannt) und *Mathematical Reasoning* (auch *Quantitative reasoning* genannt) (Geary, 1994). Beide unterschiedlichen Faktoren sollen im Folgenden von einander abgegrenzt werden: Spearman deutete schon 1927 darauf hin, dass mathematische Fähigkeiten „have much in common over and above *g*“ (Spearman, 1927, S. 251). Thurstone benannte diesen weiteren, über die generelle Intelligenz hinausgehend varianzaufklärenden Faktor in seiner Theorie der Primary Mental Abilities *Number*, die mit der Gewandtheit bei der Ausführung einfacher Rechenoperationen umschrieben wurde (Süß, 2003). Dieser gilt als einer der invariantesten Intelligenzfaktoren, ließ sich also in zahlreichen Faktorenanalysen nachweisen (vgl. auch Carroll, 1993 und Vernon, 1987). Pawlik (1971) nannte den Faktor den – neben *Verbal Comprehension (V)* – „best confirmed of all aptitude factors known“ (S. 546-547), French (1951) „the clearest of them all“ (S. 225). Carroll (1993) ordnet ihn in seinem Three-Stratum-Modell als Geschwindigkeitsfaktor (Stratum III) dem Faktor *Broad Cognitive Speediness* zu.

Neben dem Faktor *Number* oder *Numerical Facility* lässt sich ein weiterer mathematischer Faktor nachweisen, der je nach Autor *Mathematical Reasoning*, *Arithmetical Reasoning*, häufig aber auch nur *Reasoning* genannt wird. Bei diesem Faktor liegt das Hauptaugenmerk weniger auf einfachen Rechenoperationen sondern auf schlussfolgerndem Denken auf Basis numerischer Informationen. In Carrolls (1993) Modell lässt sich dieser Faktor als *level factor* ebenfalls dem

Stratum III der fluiden Intelligenz zuordnen. Er geht aber auch weiter davon aus, dass an der Beantwortung mathematischer Aufgaben noch weitere zentrale Intelligenzfaktoren beteiligt sind wie *Sequential Reasoning*, *Induction* oder *Knowledge of Mathematics*.

In Cattells (1943) beziehungsweise Horn und Cattells (1966) Intelligenztheorie lässt sich Letzteres differenzierter betrachten: Je nach der Vorwissenlastigkeit des Aufgabenmaterials laden die Testresultate stärker auf fluider oder auf kristalliner Intelligenz⁴. Aufgaben aus mathematischem Material, die eher schlussfolgerndes Denken als mathematisches Vorwissen erfordern, zeigen eine stärkere Beziehung mit g_f und vice versa (Beauducel & Kersting, 2002).

Das Ziel der hier beschriebenen Verfahrensentwicklung war es, verstärkt die fluide Intelligenz, genauer das *Mathematical Reasoning* abzubilden. Als Ergebnis der Anforderungsanalysen schien zwar auch *Knowledge of Mathematics* wichtig für den Studienerfolg zu sein, dieses erworbene Wissen lässt sich aber vergleichsweise schnell bei einem entsprechenden grundlegenden mathematischen Verständnis auffrischen. Bestimmte, auswendig gelernte Formeln, wie zum Beispiel zur Berechnung des Umfangs eines Körpers etc., sollten aus diesem Grund in keiner Aufgabe vorausgesetzt werden. Ein Itemschwerpunkt liegt auf dem Umformen und Lösen von Gleichungen und Ungleichungen (mit Brüchen und ganzen Zahlen), Operationen wie Ausklammern und Ausmultiplizieren (Distributivgesetz), die Anwendung der binomischen Formeln und das Verständnis und korrekte Anwenden von Folgerungs- und Äquivalenzumformungen (Weber, 2003). Auch diese Aufgaben sind zwar abhängig von grundlegenden Fertigkeiten, gehören jedoch bereits zum Lernstoff der Unter- beziehungsweise Mittelstufe. Bei allen verwandten Rechenregeln und Gesetzen (wie Punkt-vor-Strich-Rechnung, Kommutativgesetz, Assoziativgesetz) und Regeln zum Potenzieren und Radizieren werden Schüler nicht benachteiligt, die Mathematik in der Oberstufe abgewählt haben. Der Einfluss der Rechenfertigkeit *Numerical Facility* sollte ferner begrenzt werden, indem verwandte Zahlen ganzzahlig und maximal dreistellig sind. Entsprechend sind auch Brüche so angelegt, dass diese leicht gekürzt werden können. Zudem wurden Textaufgaben in den Test aufgenommen, die auch augenscheinlich mehrere relevante Fähigkeiten für den Berufs- und Studienalltag abdecken und somit eine höhere externe Validität, aber auch eine entsprechende Akzeptanz aufweisen. Bei diesen Aufgaben mussten die Teilnehmer in der Lage sein, aus verbalem Material formale Regeln abzuleiten.

Wesentlich für die Entwicklung der mathematischen Items war es, die besondere Zielgruppe des Self-Assessments zu berücksichtigen: Anders als bei zahlreichen Intelligenztests ist diese mit der Gruppe der Studieninteressierten klar umgrenzt: Alle Teilnehmer haben grundsätzlich eine allgemeine Hochschulreife an Gymnasien oder Gesamtschulen erworben oder streben diese an. Ein entsprechendes Niveau der Aufgaben und auch ein Umfang der verlangten Fähigkeiten musste erreicht werden. Zudem sollte auch – natürlich ohne abzuschrecken – ein kleiner

⁴ Kristalline Intelligenz „consists of discriminatory habits long established in a particular field, originally through the operation of fluid ability, but not longer requiring insightful perception for their successful operation“ (Cattell, 1943, S. 178).

Einblick in die Mathematik des späteren Studienalltags gegeben werden. So wurde verstärkt mit abstrakten Platzhaltern und weniger mit konkreten Zahlen gearbeitet. Gerade in einem Beratungssystem scheint es zudem sinnvoll, die Bedeutung der Mathematik auch in der Aufgabenauswahl und -gewichtung den Teilnehmern zu vermitteln. Nur so ist gewährleistet, dass die Studienanfänger den Stoffumfang der Mathematik und deren Wichtigkeit im Studium abschätzen können.

Im Self-Assessment der hier betrachteten Testrevision existieren 11 Items mit jeweils vier Antwortoptionen, bei denen das erste Item nicht ausgewertet wird. Dieses dient – genauso wie bei allen anderen kognitiven Tests im Self-Assessment – lediglich zur Eingewöhnung des Testteilnehmers. Die Bearbeitungszeit war auf 360 Sekunden pro Item beschränkt. Nach 300 Sekunden wurde eine Meldung eingeblendet, dass noch eine Minute verbleibt. Die entsprechende Version ist auf der Basis einer 15 Items umfassenden Vorversion anhand klassischer und probabilistischer Itemanalysen und -selektionen entstanden. Die Entwicklung der ersten Testversion wurde von Weber (2003) vorgenommen, die Weiteren vom Autor dieser Arbeit.

Die Ergebnismeldung für den Matrizen-test entspricht dem Vorgehen beim Mathematik-test.

5.3.3 Technische Realisierung

Im August 2002 wurde die erste Version, Februar 2003 die zweite, Februar 2004 die dritte und März 2004 die vierte Version des Self-Assessments für Informatik, Elektrotechnik und Technische Informatik im Internet freigeschaltet, konzipiert und programmiert durch den Autor dieser Arbeit. Technisch basiert das Beratungssystem auf einer Kombination von PHP-Skripten (primär für das Portal und das Datenhandling) und mehreren Adobe Flash⁵-Animationen (primär für die Darbietung des Beratungssystems, die Aktualisierung der Schul- und Studiennoten und die Akzeptanzbefragung). Das Self-Assessment wurde auf einem Server des Rechenzentrums der RWTH Aachen betrieben. Die Darstellung war sowohl für Windows- als auch für alternative Betriebssysteme (zum Beispiel Linux) gewährleistet. Alle wichtigen Internet-Browser (Microsoft Internet Explorer, Mozilla, Netscape Navigator, Firefox, Opera etc.) wurden unterstützt. Die Testdarbietung konnte nach einem Herunterladen des Tests offline erfolgen, so dass keine hohen Verbindungskosten durch die seinerzeit üblichen Internet-Wählverbindungen anfielen.

Ab dem Frühjahr 2003 wurde von allen Teilnehmern eine kurze Registrierung verlangt. Dabei musste eine E-Mailadresse angegeben werden, an die automatisiert ein Benutzerpasswort zugestellt wurde. Mit diesem sogenannten *opt-in*-Vorgehen ließ sich erreichen, dass von jedem

⁵ Die aufgeführten Warenzeichen und Markennamen sind Eigentum der jeweiligen Hersteller.

Teilnehmer eine valide E-Mailadresse bekannt war⁶. Mit diesem Passwort konnten sich die Interessierten zu einem selbst gewählten Zeitraum im Portal einloggen und die Beratung aufrufen. Auch ein nachträgliches Aufrufen der Feedbackseiten war damit möglich.

5.3.4 Verwandte Auswertungsverfahren

In dieser Studie werden zuerst die beschriebenen Skalen der Testbatterie mit ihren Rohwertverteilungen, den Skalenmittelwerten, Standardabweichungen und den internen Konsistenzen (erfasst mit Cronbachs- α) berichtet. Die Abweichungen der Rohwertverteilungen von einer Normalverteilung lassen sich mithilfe von Q-Q-Diagrammen darstellen. Dabei werden die beobachteten Summenwerte mit den erwarteten Normalwerten gemeinsam in einem Diagramm dargeboten (Bühl & Zöfel, 2005).

Um einen gemeinsamen kognitiven Testscore *KOGN* zu erhalten, werden die Summenscores des Mathe-, des Matrizen- und des Logiktests z-standardisiert und dann gemittelt (siehe auch Abschnitt 5.3.4). Dieses Vorgehen wird unter anderem von Ackerman und Cianciolo (2000) durchgeführt, um eine stabilere Messung der kognitiven Fähigkeiten (*g*-Faktor) zu erhalten. Yeo und Neal (2004) gehen bei Multilevel-Analysen genauso vor. Das Vorgehen entspricht dem induktiven Ansatz zur Erfassung des *g*-Faktors (Jensen, 1998), wobei der Faktor durch die Aggregation vieler stark *g*-angereicherter Tests approximiert wird. Eine direkte Messung von *g* über einen einzelnen Test birgt im direkten Vergleich zum induktiven Ansatz die Gefahr, *g* mit testspezifischer Varianz zu vermischen. Die Reliabilität des Composite-Scores wird mithilfe von Cronbachs stratifiziertem Alpha berechnet (Osburn, 1996).

Interkorrelationen der Skalen werden – unter anderem doppelt minderungskorrigiert (Lienert & Raatz, 1998) zum absoluten Vergleich – berichtet. Aufgrund der hohen Stichprobenumfänge wird auf eine Signifikanzüberprüfung der Interkorrelationen verzichtet (zur Diskussion siehe Brandstätter, 1999). Stattdessen wird die Effektgröße *d* (Cohen, 1992) berichtet, die nach einer Formel von Aaron, Kromrey und Ferron aus den Interkorrelationen berechnet wird (Vacha-Haase & Thompson, 2004). Effektstärken von $0,2 \leq d < 0,5$ werden als gering, $0,5 \leq d < 0,8$ als mittel und $d \geq 0,8$ als groß eingeschätzt (Cohen, 1992) und in der Tabelle mit ein bis drei Asterisken gekennzeichnet (siehe die jeweiligen Fußzeilen der Tabellen).

Im Hauptteil der Arbeit werden Ergebnisse nur auf Skalenebene dargestellt. Im Anhang (siehe Abschnitt 11) werden zudem die Trennschärfen und die Schwierigkeiten der einzelnen Items berichtet. Um auch bei quantitativen, nicht dichotomen Daten eine Schwierigkeitsaussage zu erhalten, wird ein modifizierter Popularitätsindex nach Dahl (1971) verwandt. Die berichteten

⁶ Auf Nachfrage wurde Schulen ein spezieller Schulzugang eingerichtet, so dass die Schüler einen verkürzten Anmeldeprozess – ohne Wartezeiten auf die Passwortmail – verwenden konnten. Mit diesem zusätzlichen Anmeldeweg wurde auf die zeitlichen Beschränkungen in Schulen reagiert. Eine valide E-Mailadresse kann bei diesem Zugangsweg nicht garantiert werden.

Trennschärpen lassen sich nach Ebel (1979) wie folgt klassifizieren: Danach sind Items mit Trennschärpen von $r_{it} < 0,20$ als „unbrauchbar“ zu bezeichnen und solche mit Koeffizienten von $0,20 \leq r_{it} < 0,30$ als „wenig brauchbar“. Items mit $0,30 \leq r_{it} < 0,40$ werden ferner als „brauchbar“ und solche mit $r_{it} \geq 0,40$ als „gut“ eingestuft.

Exploratorisch wird die Faktorenstruktur des Self-Assessments mittels einer Hauptkomponentenanalyse erfasst. Die Anzahl der zu extrahierenden Faktoren wurde dabei vorab durch Horns Parallelanalyse (1965) festgestellt. Verwandt wurde für die Berechnungen das Softwarepaket SPSS in der Version 12.0.2 (SPSS Inc., 2003) mit einer Erweiterung nach O'Connor (2000).

5.4 Ergebnisse

In die betrachtete Stichprobe aufgenommen sind alle Testdurchläufe, die laut Data-Screening (siehe Abschnitt 5.3.1) als ernsthafte Bearbeitungen klassifiziert wurden. Darüber hinaus wurde an dieser Stelle auf alle Datensätze verzichtet, die im Rahmen der Studie zur Kriteriumsvalidierung (siehe Abschnitt 7.4) erhoben worden sind, um die Testoptimierung und die Validitätsanalysen trennen zu können. Die Stichprobengröße reduziert sich durch dieses Vorgehen auf $n = 8\,633$ beziehungsweise bei dem Merkmal Handlungskontrolle (*HK*) auf $n = 6\,181$. Dabei wurde sichergestellt, dass nur identische Items einer gemeinsamen Analyse unterzogen wurden.

5.4.1 Rohwerteverteilung der Skalen

Die drei kognitiven Skalen zeigen aufgrund ihres Umfangs von jeweils 10 gewerteten Items eine mögliche Rohwerteverteilung zwischen 0 und 10 Punkten. Bei einer mittleren Schwierigkeit des Materials für die Zielstichprobe müsste theoretisch der resultierende Summenwert bei 5,0 liegen. Empirisch zeigen sich Mittelwerte von 4,751 (*MATH*), 5,154 (*MAT*) beziehungsweise 6,200 (*LOG*) (siehe Tabelle 5.1). Die Standardabweichungen der Summenwerte schwanken bei diesen drei Skalen zwischen $SD = 2,211$ (*LOG*) und 2,334 (*MAT*). Im Mittel wird die Hälfte aller kognitiven Items korrekt beantwortet (5,368). Die Skala *LOG* ist – wie die mittleren Summenwerte andeuten – für die Zielgruppe etwas leichter ausgefallen als die beiden anderen Skalen. Aber auch bei diesem Subtest deutet nichts auf einen ausgeprägten Deckeneffekt hin (alle Scoreverteilungen sind dem Anhang – Abschnitt 11.25a – zu entnehmen). Entsprechend lassen sich in den Q-Q-Diagrammen (zur Erläuterung siehe Abschnitt 5.3.4 oder Bühl & Zöfel, 2005) keine bedeutsamen Abweichungen von der angenommenen Normalverteilung erkennen (siehe Abschnitt 11.25b für *LOG*, 11.36b für *MAT* und 11.47b für *MATH*). Die volle Punktzahl haben beim Logiktest 5,5 % der Teilnehmer (absolut 475 Personen), beim Mathetest 2,4 % (204

Personen) und 3,4 % beim Matrizenstest (293 Personen) erreicht. Somit ist auch der Logiktest in der Lage, im oberen Leistungsspektrum zu diskriminieren, wenn auch etwas schlechter als die beiden anderen kognitiven Verfahren.

Betrachtet man die mittleren Skalenwerte der sechs Interessendimensionen nach Holland (zur Beschreibung der Skalen siehe Abschnitt 3.4.3), so zeigen sich große Mittelwerts- und somit Popularitätsunterschiede in der hier betrachteten Stichprobe (siehe Tabelle 5.1). Die theoretischen Rohwerteverteilungen der einzelnen Skalen verlaufen zwischen 10 und 50 Punkten. Somit liegt das theoretische Mittel bei 30 Punkten. Die Skala *INT-I* (gebildet aus Beschreibungen zu wissenschaftlich-analytischen Tätigkeiten) liegt mit einem Skalenmittelwert von $M = 35,396$ ($SD = 6,900$) über dem theoretischen Mittel und stellt zudem die „populärste“ Skala dar. Der Mittelwert der Skala *INT-E* mit Tätigkeitsbeschreibungen zum unternehmerischen Handeln ($M = 30,543$, $SD = 6,844$) liegt ebenfalls noch knapp über dem theoretischen Mittel. „Unpopulärer“ fallen die beschriebenen Tätigkeiten der Skalen *INT-C* ($M = 23,517$, $SD = 5,918$), *INT-S* ($M = 24,341$, $SD = 6,487$), *INT-A* ($M = 25,049$, $SD = 7,280$) und *INT-R* ($M = 28,028$, $SD = 7,747$) aus. Die vom theoretischen Wert abweichenden Skalenmittel gehen sowohl mit einer negativen Schiefe der Rohwerteverteilung von *INT-I* (Schiefe = $-0,254$) als auch mit einer positiven Schiefe von *INT-C* ($0,443$), *INT-A* ($0,371$) und *INT-S* ($0,303$) einher. Die Q-Q-Diagramme (siehe Abbildungen 11.2 bis 11.12) deuten aber dennoch nur auf geringe Abweichungen von einer Normalverteilung hin. Betrachtet man nicht wie bisher die Skalenmittelwerte, sondern die relative Häufigkeit der Three-Letter-Codes nach Holland (zur Erläuterung siehe Abschnitt 3.4.3), so zeigt sich, dass 11,5 % der Personen in dieser Stichprobe die Interessenstruktur *I-R-E* zeigen und weitere 8,1 % *I-E-R*. Entsprechend interessieren sich die Personen am häufigsten für wissenschaftlich-analytische, praktisch-technische und unternehmerische Tätigkeiten.

Betrachtet man die anderen nicht-kognitiven Skalen weiter, so fällt die ausgeprägte negative Schiefe bei der Skala zur Selbstwirksamkeit *SW* (siehe Grafik 11.13 oder 11.14) auf. Ein unerwünschter Deckeneffekt lässt sich ebenfalls erkennen. Die theoretische Scoreverteilung liegt zwischen 10 und 40, das theoretische Mittel somit bei 25. Insgesamt 21,2 % der Personen zeigen einen Summenwert von über 35 Punkten, beschreiben sich also im besonderen Maße als selbstwirksam, schätzen diese sich doch auf Itemebene auf der vierstufigen Skala im Mittel mit 3,5 Punkten (also genau zwischen „Stimme mit Einschränkung zu“ und „Stimme voll zu“) ein. Der Skalenmittelwert liegt bei $M = 31,694$ ($SD = 4,474$).

Wie auch schon bei der Selbstwirksamkeit zeigen sich auch die Items zur Handlungskontrolle als zu „populär“ (siehe Abbildung 11.19, 11.21 und 11.23): Die Personen präsentieren sich im besonderen Maße handlungs- und weniger lageorientiert. Die Skala *HOP* besitzt hierbei die geringste Schiefe ($-0,226$), die Skala *HOT* ($-0,703$) die ausgeprägteste. Bei der Skala *HOT* zeigen immerhin 20,5 % der Teilnehmer Summenwerte auf der vierstufigen Skala von über 35 Punkten. Die theoretische Skalenverteilung entspricht der bereits beschriebenen Selbstwirk-

Tabelle 5.1
Itemanzahl, Mittelwerte, Standardabweichung und interne Konsistenz der konstruierten Skalen

	<i>M</i>	<i>SD</i>	Skalenbereich	Cronbachs- α	Itemanzahl
Interessenfragebogen					
gesamt	166,875	22,588	60–300	0,916	60
Realistic	28,028	7,747	10–50	0,891	10
Investigative	35,396	6,900	10–50	0,869	10
Artistic	25,049	7,280	10–50	0,821	10
Social	24,341	6,487	10–50	0,834	10
Enterprising	30,543	6,844	10–50	0,850	10
Conventional	23,517	5,918	10–50	0,817	10
Selbstwirksamkeit	31,694	4,474	10–40	0,783	10
Extrinsische Lernmotivation	28,977	4,628	10–40	0,719	10
Handlungskontrolle					
gesamt	88,214	12,470	30–120	0,888	30
Tätigkeitsausführung	31,157	5,090	10–40	0,789	10
Misserfolg	29,639	4,976	10–40	0,773	10
Handlungsplanung	27,418	5,750	10–40	0,810	10
Logiktest	6,200	2,211	0–10	0,592	10
Mathematiktest	4,751	2,329	0–10	0,643	10
Matrizentest	5,154	2,334	0–10	0,681	10

Anmerkungen: $n = 8633$, für die Skalen zur Handlungskontrolle gilt $n = 6181$. M steht für den Skalenmittelwert, SD für die Standardabweichung. Die interne Konsistenz des Gesamtfragebogens zum beruflichen Interesse und der Handlungskontrolle wurde mit Cronbachs Statifiziertem Alpha berechnet.

samkeit. Die Skalenmittelwerte liegen für *HOM* bei $M = 29,638$ ($SD = 4,976$), für *HOP* bei $M = 27,418$ ($SD = 5,750$) und für *HOT* bei $M = 31,157$ ($SD = 5,090$).

Auf eine vergleichsweise geringere Abweichung von der Normalverteilung deutet die Skala der extrinsischen Lernmotivation *ELM* hin (Schiefe = $-0,191$). Dort haben insgesamt 7,7 % der Teilnehmer Punktwerte – bei derselben theoretischen Scoreverteilung wie bei *SW* – über 35 erzielt. Der Skalenmittelwert liegt bei $M = 28,977$ ($SD = 4,628$). Der Deckeneffekt ist also weniger ausgeprägt als bei den vorher beschriebenen nicht-kognitiven Skalen.

5.4.2 Interne Konsistenz der Skalen

Die interne Konsistenz der Skalen – berechnet mit Cronbachs-Alpha – schwankt zwischen $\alpha = 0,592$ beim Logiktest (*LOG*) und 0,891 bei der Skala Realistic des Interessenfragebogens (*INT-R*)⁷.

Besonders konsistent sind die Skalen des Interessenfragebogens: Cronbachs-Alpha liegt bei diesen mindestens bei $\alpha = 0,810$. *INT-R* als Skala für die praktisch-technischen Interessen deutet mit $\alpha = 0,891$ auf die höchste Konsistenz nicht nur der Interessenskalen, sondern der gesamten Testbatterie hin, die Skala *INT-C* (verwaltend-konventionell) mit $\alpha = 0,817$ auf die geringste des Interessenfragebogens. Erwartungsgemäß weisen exploratorische Faktorenanalysen (Hauptkomponentenanalysen) der einzelnen Skalen des Interessenfragebogens auf eine hohe Homogenität hin: Die jeweilig ersten Faktoren klären zwischen 38,2 % (*INT-C*) und 50,9 % (*INT-R*) der gesamten Skalenvarianz auf. Nahezu gleich gute Konsistenzwerte erzielen die Skalen der Handlungskontrolle mit $\alpha = 0,773$ bis 0,810 (Varianzaufklärung des ersten Faktors zwischen 33,2 % und 37,1 %), aber auch der Selbstwirksamkeit mit $\alpha = 0,783$ (34,1 %). Vergleichsweise geringe Werte erreicht die Skala der extrinsischen Lernmotivation *ELM* mit $\alpha = 0,719$ (29,0 %).

Geringere Konsistenzen zeigen die drei kognitiven Tests: zwischen $\alpha = 0,592$ (*LOG*) und 0,681 (*MAT*) liegen dort die Kennwerte. Anders als beim homogenen Itemmaterial der nicht-kognitiven Fragebogen zeigen sich beim heterogen entwickelten Itemmaterial der kognitiven Tests geringere, aber dennoch zufriedenstellende Varianzaufklärungen der jeweilig ersten Faktoren (Varianzaufklärung zwischen 21,6 % und 26,6 %).

Mithilfe von Cronbachs stratifiziertem Alpha lassen sich die Gesamtreliabilitäten des Interessenfragebogens ($\alpha = 0,916$), des Fragebogens zur Handlungskontrolle ($\alpha = 0,888$) und der kognitiven Tests insgesamt ($\alpha = 0,842$) berechnen.

Eine ausführliche Übersicht über die Reliabilitäten, Scoremittelwerte und Standardabweichungen bietet Tabelle 5.1.

⁷ Alle Skalen beziehungsweise Subtests umfassen einheitlich 10 Items, siehe auch Tabelle 5.1.

5.4.3 Interkorrelationen der Subskalen

Die Tabelle 5.2 gibt die Interkorrelationen der Fragebogen-Subskalen sowie der Testscores der gesamten Testbatterie wieder. Analog zum Vorgehen nach Ackerman und Cianciolo (2000) wird zusätzlich ein Composite-Score *KOGN* aus allen drei Skalen der kognitiven Studierfähigkeit als *unit-weighted z-score* gebildet (siehe auch Abschnitt 5.3.4). Außerdem werden zusätzlich zu den einzelnen Dimensionen des Interessenfragebogens die Werte des *C-Index* (Brown & Gore, 1994) berichtet, der die objektive Passung beziehungsweise die Kongruenz zu den IT-Studiengängen hinsichtlich der Interessenstruktur erfasst (siehe Abschnitt 5.3.2). Wie bereits dargelegt, wird die Interessenstruktur der Umwelt (hier Universität im Fachbereich Informatik oder Elektrotechnik) – entsprechend eines Expertenratings (Bergmann & Eder, 1992) und einer empirischen Überprüfung (Zimmerhofer, 2003) – als *I-R-E* definiert, also als wissenschaftlich-analytisch, praktisch-technisch und unternehmerisch (zur Erläuterung der Holland-Dimensionen siehe Abschnitt 3.4.3).

Auf eine Angabe der Signifikanz wird aufgrund der geringen Aussagekraft bei hohen Stichprobengrößen verzichtet. Korrelationen über $r = 0,200$ würden bei dieser Stichprobengröße und $\alpha = 0,050$ bereits signifikant werden. Stattdessen werden die Korrelationskoeffizienten in die Effektstärke d transformiert (Vacha-Haase & Thompson, 2004) und Cohens Klassifikation zu deren Einschätzung verwendet (Cohen, 1992) (siehe auch 5.3.4). Die internen Konsistenzen werden als Diagonalelemente in die Tabelle aufgenommen. Für den kognitiven Composite-Score, den Fragebogen für die Handlungskontrolle und für das Interesse wird jeweils Cronbachs stratifiziertes Alpha (Osburn, 1996) berichtet.

Die Tabelle 5.3 ermöglicht eine Abschätzung der doppelt minderungskorrigierten Interkorrelationen, um die Korrelation der „wahren“ Werte zu ermitteln. Dabei zeigt sich auf Subtestebene, dass die kognitiven, aber auch die nicht-kognitiven Skalen untereinander höher korrelieren als jeweils mit Skalen der anderen Messintention. Anders ausgedrückt sind die konvergenten Korrelationen im Mittel höher als die diskriminanten.

Dieser Umstand gilt ebenfalls für die Skala zur Selbstwirksamkeit, wenn auch das Bild dort weniger eindeutig ist: So korreliert *SW* hoch mit den nicht-kognitiven (also konvergenten) Skalen *ELM* ($r = 0,486$) und *HK* ($r = 0,619$) und geringer mit den kognitiven Skalen: So zeigen sich Korrelationen von $r = 0,140$ (*MAT*), $r = 0,149$ (*LOG*), $r = 0,218$ (*MATH*) und von $r = 0,188$ (Composite-Score *KOGN*).

Auch die Interessenskalen zeigen einen statistischen Zusammenhang sowohl mit den anderen nicht-kognitiven als auch mit den kognitiven Verfahren: Gerade die Skala *INT-I*, die wissenschaftlich-analytische Tätigkeiten beschreibt, korreliert hoch mit den kognitiven Subtests, zum Beispiel mit dem Composite-Score *KOGN* mit $r = 0,290$. Dies deutet darauf hin, dass das Interesse und die kognitiven Leistungen, die für das Ausüben benötigt werden, nicht

Tabelle 5.2
Interkorrelationen der konstruierten Skalen (nicht minderkorrigiert)

Skala	<i>C-Index</i>	<i>INT-R</i>	<i>INT-I</i>	<i>INT-A</i>	<i>INT-S</i>	<i>INT-E</i>	<i>INT-C</i>	<i>ELM</i>	<i>SW</i>	<i>HK</i>	<i>HOM</i>	<i>HOP</i>	<i>HOT</i>	<i>KOGN</i>	<i>LOG</i>	<i>MATH</i>	<i>MAT</i>
Interessenkongruenz (<i>C-Index</i>)	0,916	0,319***	0,467***	-0,199*	-0,212*	-0,459***	-0,139*	-0,037	0,104*	0,036	0,011	-0,031	0,113*	0,230**	0,170*	0,213*	0,158*
Realistic (<i>INT-R</i>)	0,319***	0,891	0,347***	-0,150*	0,009	0,084	0,211*	0,047	0,177*	0,200*	0,226**	0,105*	0,151*	0,134*	0,069	0,150*	0,097
Investigative (<i>INT-I</i>)	0,467***	0,347***	0,869	0,078	0,138*	0,190*	0,043	0,203*	0,375***	0,286**	0,184*	0,180*	0,317***	0,290**	0,240**	0,265**	0,178*
Artistic (<i>INT-A</i>)	-0,199*	-0,150*	0,078	0,821	0,370***	0,281**	0,034	0,044	0,001	-0,043	-0,053	-0,013	0,040	-0,089	-0,028	-0,121*	-0,061
Social (<i>INT-S</i>)	-0,212*	0,009	0,138*	0,370***	0,834	0,326***	0,166*	0,078	0,056	0,110*	0,266**	0,218*	0,074	-0,044	-0,022	-0,084	0,086
Enterprising (<i>INT-E</i>)	-0,459***	0,084	0,190*	0,281**	0,326***	0,850	0,333***	0,271**	0,209*	0,110*	0,227**	0,218*	0,156*	-0,101*	-0,101*	-0,120*	0,068
Conventional (<i>INT-C</i>)	-0,139*	0,211*	0,043	0,034	0,166*	0,333***	0,817	0,165*	0,055	0,017	0,017	0,218*	0,139*	-0,101*	-0,114*	-0,095	-0,074
Extrinsische Lernmot. (<i>ELM</i>)	-0,037	0,047	0,203*	0,044	0,078	0,271**	0,165*	0,719	0,364***	0,392***	0,017	0,444***	0,440***	0,022	-0,008	0,058	0,003
Selbstwirksamkeit (<i>SW</i>)	0,104*	0,177*	0,375***	0,001	0,056	0,209*	0,055	0,364***	0,783	0,513***	0,418***	0,340***	0,462***	0,152*	0,101*	0,155*	0,102*
Handlungskontrolle gesamt (<i>HK</i>)	0,036	0,200*	0,286**	-0,043	0,110*	0,277**	0,180*	0,392***	0,513***	0,888	0,672***	0,857***	0,825***	-0,015	-0,046	0,016	-0,006
Missertögl (<i>HOM</i>)	0,011	0,226**	0,184*	-0,053	0,061	0,227**	0,059	0,017	0,418***	0,672***	0,773	0,329***	0,297**	0,005	-0,001	0,007	0,006
Handlungsplanung (<i>HOP</i>)	-0,031	0,105*	0,180*	-0,013	0,121*	0,266**	0,218*	0,444***	0,340***	0,857***	0,329***	0,810	0,648***	-0,097	-0,112*	-0,048	-0,069
Tätigkeitsausführung (<i>HOT</i>)	0,113*	0,151*	0,317***	-0,040	0,074	0,156*	0,139*	0,440***	0,462***	0,825***	0,297**	0,648***	0,789	0,068	0,015	0,086	0,059
Composite-Score kognitiv (<i>KOGN</i>)	0,230**	0,134*	0,290**	-0,089	-0,044	-0,101*	-0,120*	0,022	0,152*	-0,015	0,005	-0,097	0,068	0,842	0,792***	0,790***	0,771***
Logiktest (<i>LOG</i>)	0,170*	0,069	0,240**	-0,028	-0,025	-0,068	-0,114*	-0,008	0,101*	-0,046	-0,001	-0,112*	0,015	0,792***	0,592	0,453***	0,410***
Mathematiktest (<i>MATH</i>)	0,213*	0,150*	0,265**	-0,121*	-0,022	-0,084	-0,095	0,058	0,155*	0,016	0,007	-0,048	0,086	0,790***	0,453***	0,643	0,405***
Matrizentest (<i>MAT</i>)	0,158*	0,097	0,178*	-0,061	-0,056	-0,084	-0,074	0,003	0,102*	-0,006	-0,006	-0,069	0,059	0,771***	0,410***	0,405***	0,681

Anmerkungen: Die Stichprobengrößen schwanken zwischen $6181 \leq n \leq 8633$. Die internen Konsistenzen erscheinen als Diagonalelemente und sind grau hinterlegt. Für *KOGN*, *HK* und dem *C-Index* wurden diese Werte mithilfe von Cronbachs stratifiziertem Alpha berechnet. Auf eine Angabe der Signifikanz wird verzichtet (siehe Abschnitt 5.3.4). Stattdessen wird die Effektgröße d berichtet. Ein Asterisk deutet auf eine geringe Effektstärke hin ($0,2 \leq d < 0,5$), zwei Asterisken deuten auf eine mittlere ($0,5 \leq d < 0,8$) und drei auf eine große ($d \geq 0,8$).

Tabelle 5.3
Interkorrelationen der Skalen im Self-Assessment, doppelt minderungskorrigiert

Skala	C-Index	INT-R	INT-I	INT-A	INT-S	INT-E	INT-C	ELM	SW	HK	HOM	HOP	HOT	KOGN	LOG	MATH	MAT
Interessenkongruenz (C-Index)	0,916	0,354***	0,523***	-0,229**	-0,243**	-0,520***	-0,161*	-0,045	0,123*	0,040	0,013	-0,036	0,133*	0,262**	0,231**	0,277**	0,200*
Realistic (INT-R)	0,354***	0,891	0,395***	-0,175*	0,010	0,096	0,248**	0,059	0,354***	0,225**	0,272**	0,124*	0,180*	0,155*	0,095	0,199*	0,124*
Investigative (INT-I)	0,523***	0,395***	0,869	0,092	0,162*	0,222*	0,051	0,257*	0,455***	0,325**	0,224**	0,214*	0,383***	0,339**	0,335***	0,354***	0,231**
Artistic (INT-A)	-0,229**	-0,175*	0,092	0,821	0,448***	0,336***	0,042	0,057	0,002	-0,051	-0,066	-0,016	-0,050	-0,107*	-0,040	-0,167*	-0,082
Social (INT-S)	-0,243**	0,010	0,162*	0,448***	0,834	0,387***	0,201*	0,101*	0,069	0,128*	0,076	0,148*	0,091	-0,053	-0,036	-0,030	-0,075
Enterprising (INT-E)	-0,520***	0,096	0,222*	0,336***	0,387***	0,850	0,400***	0,346***	0,256**	0,319***	0,280**	0,321***	0,191*	-0,119*	-0,096	-0,114*	-0,110*
Conventional (INT-C)	-0,161*	0,248**	0,051	0,042	0,201*	0,400***	0,817	0,215*	0,068	0,212*	0,074	0,268**	0,173*	-0,145*	-0,163*	-0,131*	-0,100*
Extrinsische Lernmot. (ELM)	-0,045	0,059	0,257**	0,057	0,101*	0,346***	0,215*	0,719	0,486***	0,490***	0,023	0,582***	0,585***	0,029	-0,013	0,086	0,040
Selbstwirksamkeit (SW)	0,123*	0,212*	0,455***	0,002	0,069	0,256**	0,068	0,486***	0,783	0,615***	0,538***	0,427***	0,588***	0,188*	0,149*	0,218*	0,140*
Handlungskontrolle gesamt (HK)	0,040	0,225**	0,325***	-0,051	0,128*	0,319***	0,212*	0,490***	0,615***	0,888	0,811***	1,000***	0,986***	-0,017	-0,063	0,021	-0,007
Misserfolg (HOM)	0,013	0,272**	0,224**	-0,066	0,076	0,280**	0,074	0,023	0,538***	0,811***	0,773	0,416***	0,360***	0,006	-0,002	0,010	0,008
Handlungsplanung (HOP)	-0,036	0,124*	0,214*	-0,016	0,148*	0,321***	0,268**	0,582***	0,427***	1,000***	0,416***	0,810	0,810***	-0,117*	-0,162*	-0,066	-0,093
Tätigkeitsausführung (HOT)	0,133*	0,180*	0,383***	-0,050	0,091	0,191*	0,173*	0,585***	0,588***	0,986***	0,380***	0,810***	0,789	0,083	0,022	0,121*	0,080
Composite-Score kognitiv (KOGN)	0,262**	0,155*	0,339***	-0,107*	-0,053	-0,119*	-0,145*	0,029	0,188*	-0,017	0,006	-0,117**	0,083	0,842	1,000***	1,000***	1,000***
Logiktest (LOG)	0,231**	0,095	0,335***	-0,040	-0,036	-0,096	-0,163*	-0,013	0,149*	-0,063	-0,006	-0,162*	0,022	1,000***	0,592	0,735***	0,645***
Mathematiktest (MATH)	0,277**	0,199*	0,354***	-0,167*	-0,030	-0,114*	-0,131*	0,086	0,218*	0,021	0,010	-0,066	0,121*	1,000***	0,735***	0,643	0,612***
Matrizentest (MAT)	0,200*	0,124*	0,231**	-0,082	-0,075	-0,110*	-0,100*	0,004	0,140*	-0,007	0,008	-0,093	0,080	1,000***	0,645***	0,612***	0,681

Anmerkungen: Die Stichprobengrößen schwanken zwischen $6181 \leq n \leq 8633$. Alle Korrelationen sind doppelt minderungskorrigiert. Korrelationen, die nach der Minderungskorrektur Werte von $r_{tt} > 1,000$ aufwiesen, wurden auf den per Definition maximalen Wert von $r_{tt} = 1,000$ gesetzt. Die Werte zur den internen Konsistenzen der Skalen erscheinen als Diagonalelemente. Für den Interessenfragebogen und den Composite-Score (KOGN) werden Cronbachs stratifizierte Alpha-Kennwerte berichtet. Auf eine Angabe der Signifikanz wird verzichtet (siehe Abschnitt 5.3.4). Stattdessen wird die Effektgröße d berichtet. Ein Asterisk deutet auf eine geringe Effektstärke hin ($0,2 \leq d < 0,5$), zwei Asterisken deuten auf eine mittlere ($0,5 \leq d < 0,8$) und drei auf eine große ($d \geq 0,8$).

orthogonale Faktoren sind. Höher fallen die Koeffizienten allerdings bei Korrelationen mit den Interessensskalen aus ($r = 0,375$ bei *SW* bzw. $r = 0,317$ bei *HOT*).

5.4.4 Reliabilität der Self-Assessment-Batterie und der Teilnehmerprofile

Wurde im Abschnitt 5.4.2 die interne Konsistenz der Untertests beschrieben, so ist ebenfalls die *Gesamtreliabilität der Testbatterie* von Interesse. Diese kann auf Basis einer Formel von Lienert und Raatz (1998) bestimmt werden, die die Reliabilität der Skalen ebenso wie deren Interkorrelationen (siehe Abschnitt 5.2) berücksichtigt. Die Reliabilität der Self-Assessments als Testbatterie beträgt demnach $r_{tt} = 0,918$.

Das Konzept der *Profilreliabilität* geht über die Reliabilität der Testbatterie hinaus, soll doch bei der Interpretation von Eignungsprofilen auch die Reliabilität von Differenzen zwischen Einzeltests berücksichtigt werden können. Die Profilreliabilität fällt nach Lienert und Raatz (1998) meist geringer aus als die Testreliabilität und wird ab $_{prof}r_{tt} = 0,500$ als ausreichend hoch eingeschätzt. Wiederum unter Berücksichtigung der bereits berichteten Interkorrelationen und der internen Konsistenzen beträgt die Profilreliabilität des Self-Assessments in der letzten Testrevision $_{prof}r_{tt} = 0,756$.

5.4.5 Faktorielle Struktur der Self-Assessment-Batterie

Zur Analyse der faktoriellen Struktur wird eine exploratorische Hauptkomponentenanalyse durchgeführt. Berücksichtigt werden dabei – zuerst außer den Interessenfragebogen – alle Summenscores der verwandten Fragebogen und Tests.

Tabelle 5.4
Komponentenmatrix der Testbatterie nach Varimax-Rotation

Skala	Faktor kognitiv	Faktor nicht-kognitiv
Selbstwirksamkeit	0,246	0,765
Extrinsische Lernmotivation	-0,031	0,705
Handlungskontrolle	-0,073	0,846
Logiktest	0,800	-0,011
Mathetest	0,783	0,092
Matrizentest	0,758	0,013

Anmerkungen: $n = 6181$

Anhand einer Parallelanalyse nach Horn (1965) sind bei der folgenden Hauptkomponentenanalyse zwei Faktoren zu extrahieren, die zudem alleinig Eigenwerte über 1 haben (2,034 und 1,668). In der Komponentenmatrix (siehe Tabelle 5.4) lässt sich ein kognitiver und ein nicht-kognitiver Faktor erkennen. Beide Faktoren zusammen klären 61,7 % der Varianz auf, wobei der kognitive Faktor unrotiert 33,9 % und der nicht-kognitive 27,8 % aufklären. Die Abbildung 5.2 visualisiert die Lage der Varimax-rotierten Faktoren im Komponentendiagramm.

Logik-, Mathe- und Matrizentests laden nahezu ausschließlich auf Faktor 1. Die Summenscores korrelieren allesamt mit mindestens $r = 0,758$ mit der kognitiven Komponente, aber unter $r = 0,100$ mit der nicht-kognitiven. Eine vergleichbare Klarheit in der Zuordnung lässt sich bei den Skalen zur extrinsischen Lernmotivation und zur Handlungskontrolle entdecken: Beide Fragebogenscores laden mit über 0,700 auf den nicht-kognitiven Faktor, und nur mit 0,031 beziehungsweise 0,073 auf den kognitiven Faktor. Die Selbstwirksamkeit zeigt bei der Zuordnung ein unklareres Bild, auch wenn sie erwartungsgemäß stärker auf den nicht-kognitiven Faktor lädt. Immerhin zeigt *SW* auch eine Korrelation von 0,246 mit dem kognitiven Faktor. Dies geht mit der Erwartung einher, dass die Selbstwirksamkeitserwartungen stärker als die anderen nicht-kognitiven Konstrukte abhängig von der Einschätzung der eigenen kognitiven Leistungsfähigkeit ist.

Fügt man noch den Interessenfragebogen in Form der Kongruenz zu Elektrotechnik beziehungsweise Informatik hinzu, so lässt sich eine hohe Korrelation des *C-Index* von $r = 0,419$ mit dem kognitiven Faktor und $r = 0,051$ mit dem nicht-kognitiven Faktor nachweisen. Das Interesse zeigt für die entsprechenden Themengebiete einen höheren statistischen Zusammenhang mit den kognitiven als mit den nicht-kognitiven Test- beziehungsweise Fragebogenwerten.

Zu klären ist noch, inwieweit sich die berichteten Faktorladungen interpretieren lassen: Nach Backhaus, Erichson, Plinke und Weiber (2006) können alle Ladungen von größer als 0,5 als substantiell angesehen werden. Guadagnoli und Velicer (1988) verweisen auf die Wichtigkeit der Stichprobengröße: Entsprechend ihrer Simulationen ist eine Stichprobe von $n = 300$ ausreichend, um Faktorenstrukturen zu interpretieren. Mithilfe ihrer veröffentlichten Formel lässt sich zudem die Differenz zwischen den Faktorladungen in der Stichprobe und in der Population quantifizieren. Für die hier berichtete Faktorenanalyse lässt sich die mittlere Differenz auf unter 0,020 schätzen. Aufgrund dieser Analyse lässt sich davon ausgehen, dass die berechneten Faktorladungen aufgrund ihrer Stabilität zweifelsfrei interpretiert werden können.

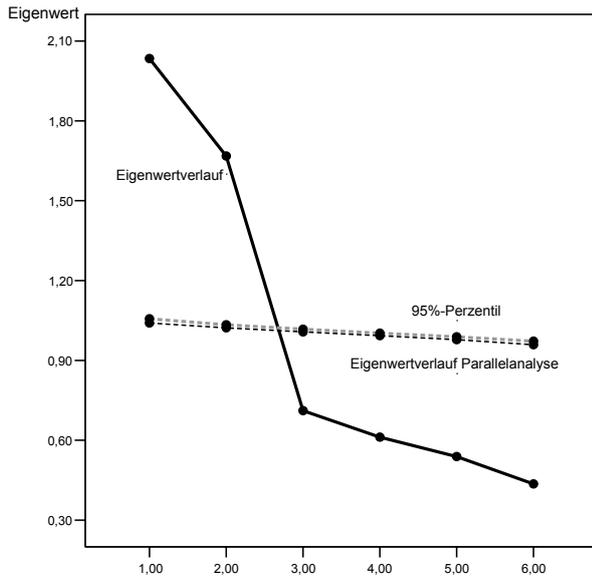


Abbildung 5.1: Screeplot der Originaldaten und der simulierten Daten im Rahmen der Parallelanalyse

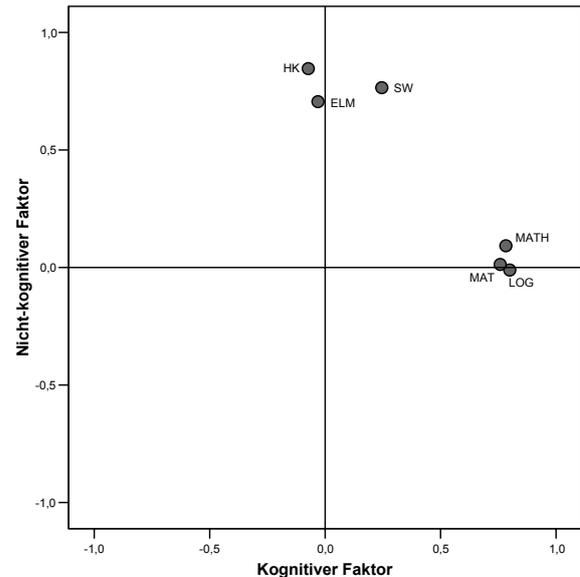


Abbildung 5.2: Ergebnis der exploratorischen Faktorenanalyse der gesamten Testbatterie

5.5 Diskussion

In der ersten Studie wurde die gesamte Testbatterie zur Studienberatung hinsichtlich der klassischen Testgütekriterien untersucht. Im Mittelpunkt standen dabei die Schwierigkeit beziehungsweise Popularität, die Trennschärfe der einzelnen kognitiven und nicht-kognitiven Items und die internen Konsistenzen der Skalen. Dabei zeigte sich, dass die internen Konsistenzen der Persönlichkeits- beziehungsweise Motivationsfragebogen (jeweils für 10-Item-Skalen) über $\alpha = 0,700$, in den meisten Fällen über $\alpha = 0,800$ liegen. Gerade der Interessenfragebogen ist inhaltlich homogen und intern konsistent. Dies ist – gerade hinsichtlich des freien Testzugangs über das Internet – ein überaus gutes Ergebnis.

Zum Vergleich dieser Kennwerte werden im Folgenden vergleichbare, publizierte Verfahren herangezogen, so der Fragebogen HAKEMP 90 von Kuhl (o.J.). Amelang und Zielinski (1997) berichten für dieses Instrument bei jeweils 12 Items Werte von $\alpha = 0,700$ bis $\alpha = 0,780$. Würden die Skalen des Self-Assessments, die jeweils nur 10 Items umfassen, durch inhaltshomogene Items entsprechend erweitert werden, so ließen sich – berechnet mithilfe der Spearman & Brown-Formel (Lienert & Raatz, 1998) – Werte von $\alpha = 0,770$ bis $\alpha = 0,810$ erreichen. Die Skalen der extrinsischen Lernmotivation und der Selbstwirksamkeit können mit Skalen aus dem Leistungsmotivationsinventar von Schuler und Proschaska (2001) verglichen werden. Bei diesem Verfahren schwanken die internen Konsistenzen zwischen $\alpha = 0,686$ und $\alpha = 0,860$ (ebenfalls 10 Items pro Skala). Die Konsistenzen der Self-Assessment-Skalen wurden mit $\alpha = 0,719$ beziehungsweise $\alpha = 0,783$ berechnet. Für den Interessenfragebogen AIST (Bergmann & Eder, 1992) werden Werte zur internen Konsistenz zwischen $\alpha = 0,810$ und $\alpha = 0,870$ (ebenfalls 10

Items pro Skala) berichtet. Ähnliche, sogar leicht höhere Werte zeigen auch die entsprechenden Skalen im Self-Assessment mit einer internen Konsistenz zwischen $\alpha = 0,817$ und $\alpha = 0,891$.

Die kognitiven Subtests erreichen unter anderem aufgrund ihrer heterogenen Inhalte geringere interne Konsistenzen. Bei einer üblichen Testlänge von 20 Items würde die interne Konsistenz dieser Skalen $\alpha = 0,744$ bis $\alpha = 0,810$ betragen. Für den I-S-T-2000R (Amthauer, Brocke, Liepmann & Beauducel, 2001) wurden Konsistenzen von $\alpha = 0,870$ bis $\alpha = 0,970$ berechnet. Blum, Hensgen und Trost (1985) berichten eine interne Konsistenz von $\alpha = 0,730$ für die Skala Mathematik in einem Studieneignungstest des Instituts für Test- und Begabungsforschung. Der Matrizen-Test Advanced Progressive Matrices von Raven (2004) erzielt bei 36 Items zwischen $\alpha = 0,830$ und $\alpha = 0,870$. Der Matrizen-Test des Self-Assessments würde bei 36 inhaltshomogenen Items einen Wert von $\alpha = 0,890$ aufzeigen. Über eine Verlängerung dieser drei kognitiven Tests mit inhaltshomogenen Items sollte nachgedacht werden, um die Messgenauigkeit der Skalen zu erhöhen. Für den Matrizen-Test sind bereits zwei zusätzliche Items entwickelt worden, die auch schon im derzeitigen Self-Assessment Verwendung finden, in dieser Arbeit aber aufgrund nicht ausreichender Stichprobengröße noch nicht evaluiert worden sind.

Exploratorisch zeigten sich in einer Hauptkomponentenanalyse erwartungsgemäß zwei leicht interpretierbare Faktoren: ein nicht-kognitiver Faktor (gebildet durch den Fragebogen zur extrinsischen Lernmotivation, zur Selbstwirksamkeit und zur Handlungskontrolle) und ein kognitiver Faktor (gebildet durch den Logiktest, durch den Mathetest und den Matrizen-Test). Eine Zwischenrolle nimmt zudem die Interessenskala ein, die sowohl auf dem kognitiven als auch auf dem nicht-kognitiven Faktor lädt. Dieser Zusammenhang sollte im Rahmen konfirmatorischer Faktorenanalysen genauer untersucht werden. Es scheint, auch wenn die Richtung des Kausalschlusses unklar bleibt, einen Zusammenhang aus IT-Interesse und IT-Fähigkeiten zu geben. Entsprechendes postuliert bereits Holland (1997) übergreifend in seiner Theorie.

Gestützt werden diese faktorenanalytischen Betrachtungen durch die berechneten Interkorrelationen der Tabelle 5.2: Die kognitiven Tests korrelieren erwartungsgemäß hoch miteinander. Alle Effektstärken liegen über $d = 0,8$. Trotz der vergleichsweise geringen Reliabilitäten liegen alle Korrelationskoeffizienten bei $r \geq 0,4$. Der Composite-Score *KOGN* hat erwartungsgemäß einen sehr hohen statistischen Zusammenhang mit den einzelnen kognitiven Skalen ($r > 0,790$). Unter den nicht-kognitiven Skalen ist die Korrelationsmatrix uneinheitlicher: *ELM*, *SW* und *HK* korrelieren mit $0,364 \leq r \leq 0,513$ ($d > 0,800$) untereinander. Der *C-Index* korreliert jedoch nicht oder nur kaum mit den anderen nicht-kognitiven Skalen. Einzelne Interessensskalen, so zum Beispiel *INT-I*, korrelieren in mittlerer Stärke sowohl mit kognitiven ($0,178 \leq r \leq 0,290$) als auch mit den nicht-kognitiven Skalen ($0,180 \leq r \leq 0,375$).

Kapitel 6

Studie 2: Profile unterschiedlicher Studienfächer

6.1 Einleitung

Ein wesentliches Ziel in der Entwicklung des webbasierten Self-Assessments zur Studienberatung war es, Studieninteressierte nicht lediglich auf die allgemeine, sondern insbesondere auf die fachspezifische Studierfähigkeit prüfen und entsprechend beraten zu können. Um diesen Zweck erreichen zu können, wurde zu Beginn des Projektes eine umfassende, fachbezogene Anforderungsanalyse (Weber, 2003; Zimmerhofer, 2003) unter anderem mit Fachvertretern der Informatik und Elektrotechnik, aber auch mit Studienberatern der Arbeitsagentur und der Hochschule durchgeführt (siehe auch Abschnitt 5.1).

Sollte es dabei gelungen sein, die Besonderheiten der beiden betrachteten Fächer zu identifizieren und bei der Testentwicklung umzusetzen, so müssten sich mit der resultierenden Testbatterie signifikante Unterschiede nicht nur zwischen den erfolgreichen und weniger erfolgreichen Studenten der Informatik und Elektrotechnik nachweisen lassen (dies wird in Abschnitt 7 näher betrachtet). Vielmehr müsste es auch möglich sein, Gruppen von Personen zu unterscheiden, die sich für unterschiedliche Studienfächer interessieren. Zahlreiche Studien (zum Beispiel Heublein & Sommer, 2002) belegen, dass Studieninteressierte im Rahmen der Selbstselektion häufig ihre Studienfachwahl aufgrund der subjektiven Eignungs- und Neigungsfeststellung vornehmen. Wenn dem auch in der hier betrachteten Zielgruppe so ist, müsste es möglich sein, Personen mit unterschiedlichen Studienpräferenzen nur aufgrund des Testprofils in eben genau diese Präferenzgruppen aufzuteilen.

Personen, die sich selbst der eigentlichen Zielgruppe der Testbatterie zuordnen – also den Studieninteressierten der Informatik, der Elektrotechnik und der Technischen Informatik – werden im Folgenden daher Interessenten gegenübergestellt, die ein geisteswissenschaftliches Studium

anstreben. Personen, die bereits ein Studium begonnen oder noch von unklaren Vorstellungen über das Studienfach ihrer Wahl berichten, werden nicht in den folgenden Analysen berücksichtigt.

Betrachtet wird außerdem, anhand welcher Prädiktoren die Trennung der genannten Gruppen am besten vorgenommen werden kann. Da gerade die eigene Interessenstruktur als wichtigstes Entscheidungskriterium für die individuelle Studienfachwahl gilt (Dawis & Lofquist, 1984; Lewin, Heublein, Schreiber, Spangenberg & Sommer, 2002; Heublein & Sommer, 2002, siehe auch Abschnitt 3.4.3), müssten sich beim Interessenfragebogen die bedeutendsten Mittelwertsunterschiede zwischen den Interessenten der Geisteswissenschaften und der Informatik/Elektrotechnik aufzeigen lassen. Ein solches Ergebnis setzt allerdings voraus, dass eine hohe Ähnlichkeit gemessener¹ und artikulierter² Interessen existiert (siehe ebenfalls Abschnitt 3.4.3). Neben dem bereits beschriebenen Interessenfragebogen werden die Fähigkeit zur Diskriminierung der übrigen nicht-kognitiven Fragebogen, der kognitiven Tests, aber auch der Abiturdurchschnittsnote miteinander verglichen. Auf Basis einer Befragung von Heublein und Sommer (2002) ist anzunehmen, dass neben der subjektiven Feststellung der Neigung insbesondere die Begabung eine entscheidende Rolle spielt. Zu vermuten ist weiterhin, dass die Abiturdurchschnittsnote in Fächern, die nicht zulassungsbeschränkt sind, keine Trennung in die unterschiedlichen Präferenzgruppen ermöglicht.

6.2 Fragestellung

In der zweiten Studie werden folgende Fragestellungen genauer betrachtet:

Es wird angenommen, dass das mittlere Test-Profil der Zielgruppe (Studieninteressierte der Informatik oder Elektrotechnik) von dem mittleren Profil der Personen, die eine Geisteswissenschaft studieren wollen, signifikant verschieden ist. Dies kann durch die Selbstselektion der Studieninteressierten erklärt werden (siehe Abschnitt 2.4), die sich entsprechend eigener Fähigkeiten, Fertigkeiten und Kenntnisse für ein Studienfach entscheiden. Ein solches Ergebnis würde die *gravitational theory* (Wilk, Desmarais & Sackett, 1995), die annimmt, dass ein Mensch auf seinem Karriereweg immer stärker Berufe und Arbeitsplätze wählt, die der eigenen Eignung und Neigung entsprechen, bereits für den ersten Schritt in eine Berufskarriere stützen. Empirisch würden sich signifikante Mittelwertsunterschiede und entsprechend hohe Effektstärken beim Vergleich der beiden Präferenzgruppen zeigen. Dies soll auch dann gelten, wenn Geschlechterunterschiede und verschiedene Geschlechterverhältnisse bei den Studienrichtungen kontrolliert werden: Auf Basis der Auskünfte des Statistischen Bundesamtes (2005a) ist festzustellen, dass die Studienfächer Informatik beziehungsweise Elektrotechnik von Männern bevorzugt werden

¹ erfasst mit einem Interessenfragebogen

² operationalisiert über eine direkte Frage nach dem präferierten Studienfach

und sich dies auch in dem Geschlechterverhältnis der hier beschriebenen Stichprobe widerspiegelt. Bei den Geisteswissenschaften ist es hingegen – wenn auch nicht so extrem ausgeprägt – umgekehrt. Allein durch diesen Umstand könnten Mittelwertsunterschiede erklärt werden, unterschiedliche Eignungs- und Neigungsprofile der Geschlechter vorausgesetzt (zum Beispiel wie in Benbow & Stanley, 1983, bezüglich der Mathematik oder in Bergmann, 2003, bezüglich der Interessen nachgewiesen). Aus diesem Grund wird hier angenommen, dass zwar Geschlechterunterschiede in den Fragebogen- und Testscores im Mittel bestehen (und somit ein Haupteffekt *Geschlecht* existiert), zugleich aber auch ein Haupteffekt *Studienfachwahl* nachweisbar ist. Ferner wird eine Interaktion der Studienpräferenz und des Geschlechts vermutet: In der Gruppe der Informatik- bzw. Elektrotechnik-Interessierten müsste demnach aufgrund der besonderen Bedeutung der erfassten Dimensionen die Geschlechterunterschiede geringer als bei der Präferenzgruppe Geisteswissenschaft ausfallen.

Weiterhin wird vermutet, dass die Testbatterie des Self-Assessments mithilfe einer logistischen Regression in der Lage ist, die Zielgruppe (die Studieninteressenten der Informatik beziehungsweise der Elektrotechnik) von der Gruppe der Studieninteressierten der Geisteswissenschaften zu trennen. Dabei sollten sich die nicht-kognitiven Prädiktoren (insbesondere das Interesse) besser geeignet zur Vorhersage der Studienpräferenz zeigen als die kognitiven. Diese Vermutung lässt sich stützen durch die Umfrageergebnisse der HIS (zum Beispiel Heublein & Sommer, 2002). Auch müsste die gesamte Testbatterie des Self-Assessments die unterschiedlichen Teilnehmergruppen besser trennen als dies die Abiturdurchschnittsnote erlaubt. Eine Selbstselektion auf Basis der Abiturnote kann bei nicht zentral zulassungsbeschränkten Fächern aufgrund der bisherigen Forschung nicht angenommen werden.

In den vorigen Forschungsfragen wurde stets angenommen, dass Unterschiede zwischen den Studienpräferenzen Informatik/Elektrotechnik und Geisteswissenschaften existieren. Still-schweigend wurde damit ebenfalls postuliert, dass keine Unterschiede in den Test-Profilen zwischen Informatik und Elektrotechnik aufzufinden sind. Gerade im Grundstudium sind beide Fächer durch einen hohen Anteil an Mathematik gekennzeichnet, so dass in der Anforderungsanalyse an der RWTH Aachen keine nennenswerten Unterschiede erarbeitet werden konnten. Eine zentrale Hypothese für alle weiteren Berechnungen ist daher, dass Analysen der Höhen und der Gestalt der Testbatterie-Profile, aber auch ein Vergleich der Skalen-Mittelwerte nicht auf einen bedeutsamen Unterschied zwischen Elektrotechnik und Informatik hinweisen. Es wird – analog zu der obigen Forschungsfrage – weiter angenommen, dass die Self-Assessment-Testbatterie nicht in der Lage ist, die Zielgruppe des Verfahrens, also die Studieninteressierten der Informatik und Elektrotechnik, anhand ihrer Testergebnisse zu separieren. Die Studieninteressierten der Technischen Informatik werden hier der Elektrotechnik zugeschlagen, da das Grundstudium dem der Elektrotechnik gleicht.

6.3 Methoden

In der zweiten Studie werden Profilunterschiede zwischen zwei Stichproben, gebildet aus Interessentengruppen unterschiedlicher Studienfächer, analysiert (siehe auch Abschnitt 6.2). Die Teilnehmergruppen werden auf Basis einer Selbstauskunft aufgeteilt, in der erfragt wurde, welche Studienfächer die Interessierten in die engere Wahl nehmen.

6.3.1 Stichprobe

Genauso wie in der Studie 1 wird auch in der Studie 2 auf den gesamten Datensatz ernsthafter Bearbeitungen zurückgegriffen. Dabei werden ausschließlich Ergebnisse von Studieninteressierten der Fächer Informatik, Elektrotechnik und Technischen Informatik beziehungsweise der Geisteswissenschaften näher ausgewertet, um die unterschiedlichen Studienprofile miteinander vergleichen zu können. Eine direkte Frage nach dem Studienwunsch wurde dabei ausgewertet. Daten von Personen, die andere Fächer studieren möchten, wurden eliminiert. Da zudem Extremgruppen gebildet werden sollten, sind nur Personen mit einer klaren Präferenz in der weiteren Analyse belassen worden. Personen, die eventuell Informatik oder ein geisteswissenschaftliches Fach studieren möchten, sind nicht weiter berücksichtigt worden. Daten von Personen, die bereits ein Studium begonnen haben, werden in den Analysen nicht weiter berücksichtigt.

Insgesamt stehen nach den oben genannten Kriterien 2 562 Studieninteressierte der Informatik, Elektrotechnik und Technischen Informatik für eine Analyse zur Verfügung. Allerdings haben nur 1 582 davon die letzte Testrevision der Skala Handlungskontrolle durchgeführt. Bei den Geisteswissenschaften haben 354 beziehungsweise 298 (Stichprobengröße der Handlungskontrolle) Interessenten das Verfahren absolviert. Im Median waren Teilnehmer beider Gruppen 19 Jahre alt ($M = 19,7$ und $SD = 2,505$ bei Informatik/Elektrotechnik beziehungsweise $M = 19,3$ und $SD = 3,037$ bei Geisteswissenschaften). In der Informatik- und Elektrotechnikgruppe waren 11 % weibliche Teilnehmer, in der geisteswissenschaftlichen Gruppe hingegen 69 %. Aufgrund der Unterschiede in der Geschlechterverteilung in den beiden Gruppen werden weitere Analysen notwendig sein.

6.3.2 Verwandte Testverfahren

Die Beschreibung der verwandten Verfahren sind dem Abschnitt 5.3.2 zu entnehmen.

6.3.3 Technische Realisierung

Die technische Realisierung ist identisch mit dem in der Studie 1 (siehe Abschnitt 5.3.3) beschriebenen Vorgehen.

6.3.4 Verwandte Auswertungsverfahren

Der Extremgruppenvergleich zwischen den Studieninteressierten der Informatik und Elektrotechnik beziehungsweise der Geisteswissenschaften wird mithilfe von drei Auswertungsverfahren untersucht: Entsprechend den Empfehlungen der US-amerikanischen Psychologengemeinschaft (American Psychological Association, 2001; Vacha-Haase & Thompson, 2004) werden zur Einschätzung der Mittelwertsunterschiede der Extremgruppen Effektstärken berichtet. Zurückgegriffen wird hier auf Cohens d (Cohen, 1992), das im Vergleich zu unterschiedlichen Effektstärken geringere Über- beziehungsweise Unterschätzungen (Bias) (Roberts & Henson, 2002) aufweist. Die gerade bei Signifikanztests hohe Abhängigkeit von den betrachteten Stichprobengrößen ist bei diesem Vorgehen nicht gegeben.

Im zweiten Schritt wird eine logistische Regression durchgeführt und die Varianzaufklärung als Nagelkerkes- R^2 berechnet (Backhaus, Erichson, Plinke & Weiber, 2006). Dabei wird kalkuliert, mit welcher Wahrscheinlichkeit eine Person mit einem bestimmten Ergebnisprofil ein Studium der Informatik, Elektrotechnik und Technischen Informatik beziehungsweise der Geisteswissenschaften aufnehmen möchte. Zentral ist zudem die Fragestellung, welche Subtest- beziehungsweise Skalenergebnisse in der Lage sind, die entsprechende Studienpräferenz am trennschärfsten vorherzusagen.

Im dritten Schritt werden so genannte *Receiver Operating Characteristic*- oder auch *Relative Operating Characteristic*-Kurven (kurz ROC-Kurven) (Bühl & Zöfel, 2005; Humphreys & Swets, 1991) verwandt, die die Prognosefähigkeiten der kognitiven und der nicht-kognitiven Testkomponenten im direkten Vergleich veranschaulichen. Die derartige Darstellung der Validität von Test- und Fragebogenscores in Bezug auf ein dichotomes Kriterium hat den wesentlichen Vorteil, dass zur Analyse der Vorhersagekraft unterschiedliche Schwellenwerte der Scores zur Dichotomisierung festgelegt werden. Damit lässt sich die Diskriminierungsfähigkeit von mehreren Prädiktoren direkt über das gesamte Kontinuum vergleichen. Die Einsatzmöglichkeiten für die Eignungsdiagnostik sind Humphreys und Swets (1991) zu entnehmen.

Im Rahmen der ROC-Analysen wird das zu erkennende Signal als *Interesse an einem Studium der Informatik oder Elektrotechnik* verstanden. Die Nullhypothese (H_0) besagt, dass kein Interesse (beziehungsweise kein Signal) herrscht, die Alternativhypothese (H_1) hingegen, dass Interesse (beziehungsweise das Signal) vorhanden ist.

Entsprechend der Signalentdeckungstheorie können folgende vier Gruppen der Entscheidung betrachtet werden:

- Personen werden korrekterweise als nicht für Informatik/Elektrotechnik interessiert eingeschätzt (richtig-negativ)
- Personen werden fälschlicherweise als nicht für Informatik/Elektrotechnik interessiert eingestuft (falsch-negativ)
- Personen werden korrekterweise als für Informatik/Elektrotechnik interessiert eingeschätzt (richtig-positiv)
- Personen werden fälschlicherweise als für Informatik/Elektrotechnik interessiert eingeschätzt (falsch-positiv)

Die Sensitivität eines Tests ist definiert als Anteil der richtig erkannt Interessierten an Informatik/Elektrotechnik (*richtig-positiv*) an der Gesamtzahl der erkannt (*richtig-positiv*) oder unerkannt Interessierten (*falsch-negativ*) (siehe Formel 6.1).

$$\text{Sensitivität} = \frac{\text{Anzahl}(\text{richtig-positiv})}{\text{Anzahl}(\text{richtig-positiv}) + \text{Anzahl}(\text{falsch-negativ})} \quad (6.1)$$

Die Spezifität eines Tests entspricht dem Anteil der erkannt Uninteressierten (*richtig-negativ*) unter den als erkannt (*richtig-negativ*) oder unerkannt uninteressierten Personen (*falsch-positiv*) (siehe Formel 6.2).

$$\text{Spezifität} = \frac{\text{Anzahl}(\text{richtig-negativ})}{\text{Anzahl}(\text{richtig-negativ}) + \text{Anzahl}(\text{falsch-positiv})} \quad (6.2)$$

Bei der ROC-Kurve wird die Sensitivität gegen den Komplementärwert der Spezifität zu 1 aufgetragen. Dies entspricht der Wahrscheinlichkeit für eine Falsch-Positive-Entscheidung, berechnet als Anzahl unerkannt uninteressierter Personen (*falsch-positiv*) unter den erkannt (*richtig-negativ*) oder unerkannt Uninteressierten (*falsch-positiv*) (siehe Formel 6.3).

$$1 - \text{Spezifität} = \frac{\text{Anzahl}(\text{falsch-positiv})}{\text{Anzahl}(\text{richtig-negativ}) + \text{Anzahl}(\text{falsch-positiv})} \quad (6.3)$$

Die Grundlage dieser ROC-Kurven wird mit der bereits beschriebenen logistischen Regression gelegt. Deren Regressionsgerade wird verwandt, um den Prädiktorenscore zu bestimmen.

Die Berechnung der Regressionsgerade und die Visualisierung der *ROC*-Kurven wird mithilfe von SPSS 12.0.2 (SPSS Inc., 2003) vorgenommen.

6.4 Ergebnisse

Im ersten Schritt wird ein Mittelwertvergleich zwischen beiden beschriebenen Präferenzgruppen vorgenommen (siehe Tabelle 6.1). Die Stichprobengröße schwankt auf der Seite der eigentlichen Zielgruppe des Beratungssystems zwischen $1582 \leq n \leq 2562$, bei der geisteswissenschaftlichen Gruppe zwischen $298 \leq n \leq 354$ ³.

Betrachtet man die resultierenden Effektstärken genauer, so fallen zwei auf, die entsprechend der Klassifikation nach Cohen (1988) als *groß* zu bezeichnen sind: Beim Mittelwertvergleich der Interessenkongruenz (*C-Index* nach Brown & Gore, 1994, siehe Abschnitt 3.4.3 und Abschnitt 5.3.2) ergibt sich eine Effektstärke von $d = 0,993$: Personen mit einem präferierten Studienfach Informatik oder Elektrotechnik gleichen somit stärker als Interessenten der Geisteswissenschaften derjenigen Umwelt, die mittels Expertenrating für Studenten der Informatik und Elektrotechnik festgestellt wurde (siehe Abschnitt 5.3.2). Die meisten Interessenten (18,9 % der Personen) der Informatik und auch der Elektrotechnik zeigen die Interessenstruktur *I* (wissenschaftlich-analytisch) *R* (praktisch-technisch) *E* (unternehmerisch), die auch im Rahmen des Expertenratings bestimmt wurde (Bergmann & Eder, 1992). Insgesamt 30,6 % der Three-Letter-Codes entfallen gemeinsam auf *I-R-E* – als häufigste Interessenstruktur – und *I-E-R* – als zweithäufigste. In der Gruppe der geisteswissenschaftlich Interessierten hingegen zeigen lediglich 0,5 % der Personen diese Three-Letter-Codes. Am häufigsten treten in dieser Gruppe die Kombinationen der Dimensionen *A* (künstlerisch-gestalterisch) *I* *E* auf: 35,1 % der geisteswissenschaftlich Interessierten präferieren die Auseinandersetzung mit diesen entsprechenden Tätigkeiten.

Eine ebenfalls bedeutsame Effektstärke lässt sich beim Composite-Score der kognitiven Testbestandteile *KOGN* (siehe Abschnitt 5.3.4) nachweisen ($d = 0,883$), insbesondere durch die kognitive Skala *MATH* ($d = 0,834$) hervorgerufen. Dieses Ergebnis deutet auf bedeutsame Mittelwertsunterschiede im Bereich der kognitiven Fähigkeiten hin: Sowohl im Composite-Score *KOGN* als auch im Summenscore *MATH* zeigt sich die eigentliche Zielgruppe jeweils um etwa eine Standardabweichung leistungsstärker als die Gruppe der geisteswissenschaftlich Interessierten.

Mittlere Effektgrößen über $d = 0,5$ lassen sich bei den Skalen *MAT* ($d = 0,604$), *LOG* (0,584) und *HK* (0,514) aufzeigen. Wie schon die Werte des Composite-Scores vermuten lassen, zeigt die eigentliche Zielgruppe im Mittel auch im Logik- und Matrizen-test bessere Ergebnisse. Zudem sind Personen der Zielgruppe stärker handlungsorientiert und schwächer lageorientiert (siehe Abschnitt 3.4.6 beziehungsweise 5.3.2). Geringe Mittelwertsunterschiede lassen sich bei Fragebogen *SW* ($d = 0,245$) nachweisen: Interessenten der Informatik/Elektrotechnik scheinen

³ Aufgrund von unterschiedlichen Revisionen des Itemmaterials schwankt die Stichprobengröße zwischen den betrachteten Subtests. Es ist sichergestellt, dass nur dann Testrevisionen zusammen ausgewertet werden, wenn diese dasselbe Itemmaterial umfassen.

Tabelle 6.1
Vergleich der Testergebnisse von Studieninteressierten der Geisteswissenschaften und der Informatik, Elektrotechnik und Technische Informatik

Gruppe		Abiturnote	C-Index	ELM	HK	SW	KOGN	LOG	MAT	MATH
Informatik,	<i>M</i>	2,363	12,617	28,927	90,982	31,754	0,109	6,424	5,433	5,001
	<i>SD</i>	0,629	4,354	4,426	11,319	4,609	0,775	2,170	2,390	2,313
Technische Informatik	<i>n</i>	2211	2562	2562	1582	2562	2562	2562	2562	2562
	<i>M</i>	2,398	8,455	29,528	84,721	30,650	-0,513	5,189	4,102	3,282
Geistes-	<i>SD</i>	0,647	4,019	4,727	12,952	4,396	0,626	2,059	2,004	1,773
	<i>n</i>	345	354	354	298	354	354	354	354	354
Effektstärke	<i>d</i>	0,055	0,993***	0,131	0,514**	0,245*	0,883***	0,584**	0,604**	0,834***

Anmerkungen: Alle Mittelwertsunterschiede bis auf *Abiturdurchschnittsnote* sind bei $\alpha = 5\%$ signifikant. Ein Asterisk deutet auf eine geringe Effektstärke hin ($0,2 \leq d < 0,5$), zwei Asterisken deutet auf eine mittlere ($0,5 \leq d < 0,8$) und drei auf eine grobe ($d \geq 0,8$).

Tabelle 6.2
Vergleich der Testergebnisse von Studieninteressierten der Elektrotechnik und der Informatik

Gruppe		Abiturnote	C-Index	ELM	HK	SW	KOGN	LOG	MAT	MATH
Informatik	<i>M</i>	2,333	11,646	29,247	89,602	32,740	0,005	6,292	4,699	5,141
	<i>SD</i>	0,653	4,405	4,577	11,613	4,011	0,780	2,236	2,305	2,342
	<i>n</i>	915	941	941	686	941	941	941	941	941
	<i>M</i>	2,266	11,744	28,748	91,033	30,849	0,138	6,477	5,011	5,568
Elektrotechnik	<i>SD</i>	0,621	4,511	4,321	11,655	4,833	0,730	2,068	2,241	2,220
	<i>n</i>	597	750	750	418	750	750	750	750	750
Effektstärke	<i>d</i>	0,106	0,020	0,112	0,123	0,426*	0,121	0,088	0,136	0,188

Anmerkungen: Alle Mittelwertsunterschiede bis auf *INT* und *LOG* sind bei $\alpha = 5\%$ signifikant. Ein Asterisk deutet auf eine geringe Effektstärke hin ($0,2 \leq d < 0,5$), zwei Asterisken deutet auf eine mittlere ($0,5 \leq d < 0,8$) und drei auf eine grobe ($d \geq 0,8$).

eine höhere Selbstwirksamkeit zu haben. Tendenziell ($d = 0,131$) zeigt sich zudem die Gruppe der Geisteswissenschaftler extrinsisch lernmotivierter. Alle bis hierhin berichteten Mittelwertsunterschiede sind – berechnet mit einem t-Test für unabhängige Stichproben – bei $\alpha = 0,050$ signifikant. Der geringste und damit auch nicht signifikante Mittelwertsunterschied zeigt sich bei der Abiturdurchschnittsnote ($d = 0,055$).

Die Tabelle 6.3 veranschaulicht die mittleren Summenscores im Self-Assessment sowohl der beschriebenen Präferenzgruppen als auch der beiden Geschlechter. Damit lässt sich überprüfen, ob die bisher berichteten Unterschiede in den Summenscores nicht lediglich auf unterschiedliche Geschlechterverhältnisse⁴ in den betrachteten Präferenzgruppen und den bereits in der Literatur häufig berichteten Fähigkeits- und Motivationsunterschieden von Männern und Frauen beruhen (zum Beispiel wie in Benbow & Stanley, 1983, bezüglich der Mathematik oder in Bergmann, 2003, bezüglich der Interessentests nachgewiesen).

Die mittleren Effektstärken für den Vergleich der männlichen und der weiblichen Teilnehmer betragen in der Präferenzgruppe *Informatik* $d = 0,149$ und in der Gruppe *Geisteswissenschaft* $d = 0,326$. Vergleicht man hingegen männliche Teilnehmer der Präferenzgruppen *Geisteswissenschaften* und *Informatik* miteinander und getrennt davon die weiblichen Teilnehmer der beiden Präferenzgruppen, so zeigen sich weitaus höhere Effektstärken: Bei einem Vergleich der weiblichen Präferenzgruppe über die Studienfächer hinweg beträgt die Effektstärke $d = 0,649$. Bei den Männern liegt die Effektstärke bei $d = 0,505$. Dies deutet daraufhin, dass es zwar Geschlechterunterschiede innerhalb der Präferenzgruppen gibt, diese aber geringer ausfallen als die Unterschiede, die zwischen den Präferenzgruppen bei einer getrennten Betrachtung der Geschlechter existieren.

Wieder fallen bei Vergleichen durch die Interessenfragebogen große Unterschiede auf: Innerhalb der Präferenzgruppen zeigen sich keine bedeutsamen Differenzen ($d = 0,010$ bzw. $d = 0,224$). Hingegen lassen sich Effektstärken von $d = 0,843$ bzw. $d = 1,111$ berechnen, wenn man dieselben Geschlechter aus unterschiedlichen Präferenzgruppen vergleicht.

Im Rahmen einer 2-faktoriellen Varianzanalyse mit den Faktoren Geschlecht und Studienpräferenz lassen sich zwei signifikante Haupteffekte (bei $\alpha = 0,05$) nachweisen: Sowohl das Geschlecht als auch die Studienfach-Präferenz haben einen bedeutsamen Einfluss auf die Testergebnisse. Dabei zeigen sich größere Effektstärken beim Vergleich der Präferenzgruppen als beim Vergleich der Geschlechter. Im Mittel haben Personen in der Präferenzgruppe Informatik, Elektrotechnik und Technische Informatik bessere Werte in den kognitiven Tests als die Personen in der Präferenzgruppe Geisteswissenschaften (siehe Abbildungen 6.2 bis 6.7).

⁴ Der Anteil weiblicher Teilnehmer in der Präferenzgruppe Informatik, Elektrotechnik und Technische Informatik beträgt nur 11 %, in der Gruppe der geisteswissenschaftlich Interessierten hingegen 69 % (siehe auch 6.3.1)

Tabelle 6.3
Vergleich der Testergebnisse von Studieninteressierten beider Geschlechter

		<i>C-Index</i>	<i>ELM</i>	<i>HK</i>	<i>SW</i>	<i>KOGN</i>	<i>LOG</i>	<i>MATH</i>	<i>MAT</i>
Informatik, Elektrotechnik, Technische Informatik	weiblich	<i>M</i> 12,719 <i>SD</i> 4,323 <i>n</i> 224	30,924 4,335 224	91,893 13,097 169	32,754 3,677 224	-0,043 0,761 224	6,156 2,198 224	4,853 2,333 224	4,795 2,264 224
	männlich	<i>M</i> 12,762 <i>SD</i> 4,317 <i>n</i> 1823	29,421 4,422 1823	90,814 11,134 1379	33,184 3,730 1823	0,103 0,772 1823	6,329 2,161 1823	4,999 2,294 1823	5,493 2,356 1823
Geistes- wissenschaften	Geschlechter- vergleich	<i>d</i> 0,010	0,343 *	0,089	0,116	0,190	0,079	0,063	0,302 *
	weiblich	<i>M</i> 8,214 <i>SD</i> 3,760 <i>n</i> 234	30,368 4,375 234	85,201 13,019 204	30,718 4,330 234	-0,638 0,531 234	4,855 1,971 234	3,120 1,543 234	3,744 1,704 234
Geistes- wissenschaften	männlich	<i>M</i> 9,121 <i>SD</i> 4,326 <i>n</i> 107	28,019 5,101 107	83,538 12,954 91	30,869 4,543 107	-0,291 0,722 107	5,757 2,073 107	3,514 2,156 107	4,822 2,378 107
	Geschlechter- vergleich	<i>d</i> 0,224 *	0,494 *	0,128	0,034	0,548 **	0,446 *	0,210 *	0,521 ***
Vergleich Präferenzgruppen Frauen	<i>d</i> 1,112 ***	0,128	0,512 **	0,507 **	0,907 ***	0,623 **	0,876 ***	0,525 **	
Vergleich Präferenzgruppen Männer	<i>d</i> 0,843 ***	0,294 *	0,602 **	0,557 **	0,527 **	0,270 *	0,667 **	0,283 *	

Anmerkungen: In diese Analyse wurden nur Datensätze aufgenommen, zu denen Informationen zum Geschlecht des Teilnehmers beziehungsweise der Teilnehmerin vorlagen. Ein Asterisk deutet auf eine geringe Effektstärke hin ($0,2 \leq d < 0,5$), zwei Asterisken deuten auf eine mittlere ($0,5 \leq d < 0,8$) und drei auf eine große ($d \geq 0,8$).

Neben diesen beiden Haupteffekten ist auch die Interaktion *Geschlecht*Präferenz-Gruppe* signifikant. Besonders augenscheinlich ist diese bei den kognitiven Tests und beim Interessenfragebogen (siehe 6.4 für den Interessenfragebogen, Abbildung 6.6 für den Logiktest und Abbildung 6.7 für den Mathematiktest; dort werden jeweils die mittleren Rohwerte der Teilnehmergruppen abgetragen). In der Präferenz-Gruppe zeigt sich zum Beispiel kein Unterschied zwischen der Testleistung der männlichen und der weiblichen Teilnehmer beim *Logiktest*. Dieser ist aber umso größer in der Präferenz-Gruppe der *Geisteswissenschaften*: Dort haben die weiblichen Teilnehmer signifikant schwächer abgeschnitten als die männlichen Teilnehmer. Diese haben aber wiederum schwächer abgeschnitten als die männlichen Teilnehmer in der Präferenz-Gruppe *Informatik*. Ein ähnliches Phänomen, wenn auch nicht so ausgeprägt, lässt sich noch beim Mathetest und auch beim Interessentest nachweisen. Dabei zeigt sich zwar, dass die männlichen Teilnehmer der Präferenz-Gruppe *Geisteswissenschaften* eine höhere Passung zur Umwelt *Informatikstudium* aufweisen als die weiblichen Teilnehmer. Diese Differenz existiert aber bei der alleinigen Betrachtung der Präferenz-Gruppe *Informatik* nicht mehr. Dort gleichen sich die Teilnehmer beider Geschlechter weitgehend.

Mithilfe einer linearen Kombination der Abiturdurchschnittsnote, der Werte in den kognitiven Tests und in den nicht-kognitiven Fragebogen des Self-Assessments wird im Folgenden versucht, post-hoc diese Klassifikation aller Teilnehmer in die beiden Präferenzgruppen vorzunehmen (siehe auch Tabelle 6.4). Damit wird es möglich, nicht nur die einzelnen Bestandteile des Self-Assessments getrennt auszuwerten, sondern eine Gewichtung der möglichen Prädiktoren in einer logistischen Regressionsgeraden zu betrachten, die Varianzaufklärung der einzelnen und aller Prädiktoren gemeinsam zu bewerten. Bei diesem Vorgehen zeigt sich, dass alleine die Interessenkongruenz eine Varianzaufklärung nach Nagelkerkes von $R^2 = 0,174$ ($-2LL = 1\,443,245$, $\chi^2 = 200,597$, $df = 1$, $p < 0,001$) erreicht, die sonstigen nicht-kognitiven Prädiktoren eine Aufklärung von $R^2 = 0,125$ ($-2LL = 1\,501,000$, $\chi^2 = 142,469$, $df = 5$, $p < 0,001$) und die kognitiven Prädiktoren von $R^2 = 0,153$ ($-2LL = 1\,467,904$, $\chi^2 = 175,593$, $df = 3$, $p < 0,001$). Diese drei Prädiktoren klären somit signifikant Varianz in der Gruppenaufteilung auf und ermöglichen so eine Klassifikation. Die Abiturdurchschnittsnote trägt hingegen nicht zur Varianzaufklärung ($R^2 = 0,000$, $-2LL = 1\,643,259$, $\chi^2 = 0,238$, $df=1$, $p = 0,626$) bei und wird daher hier nicht mehr weiter betrachtet.

Mit einer hierarchischen logistischen Regression wird zudem die kombinierte Varianzaufklärung berechnet. Die weiteren nicht-kognitiven Fragebogen klären über die Interessenkongruenz hinaus (Nagelkerkes- $R^2 = 0,174$) weitere Varianz auf (Δ Nagelkerkes- $R^2 = 0,090$), als auch die kognitiven über die Gesamtheit der nicht-kognitiven Verfahren ($\Delta R^2 = 0,067$) (siehe Tabelle 6.4). Insgesamt werden 33,1 % der Varianz in der Klassenzuordnung durch das Self-Assessment erklärt. Nach der Klassifikation von Backhaus, Erichson, Plinke und Weiber (2006) ist die Höhe der Varianzaufklärung *akzeptabel* bis *gut*. Bei einem Trennwert von 0,500 würden 87,1 % korrekt klassifiziert werden.

Bei einer reinen Zufallszuweisung würde die Fehlerquote statt bei 12,9 % bei 26,7 % liegen⁵. Nach einem Angleichen der Stichprobengrößen auf jeweils $n = 298$ steigt die Fehlerrate bei Zufallszuordnungen auf 50,0 %. Die Zuweisung auf Basis der logistischen Regression würde in 77,0 % der Fälle korrekt erfolgen.

Die resultierende Regressionsgerade zur Prädiktion der Gruppenzugehörigkeit auf Basis der Abiturdurchschnittsnote, der Test- und Fragebogenergebnisse wird im Folgenden nochmals verwendet. Die auf dieser Grundlage erstellten *Receiver Operating Characteristic*-Kurven (oder auch *ROC-Kurven*) ermöglichen, die Validität der Prädiktorguppen – die kognitiven Tests, die nicht-kognitiven Fragebogen und das Abitur – differenziert über das gesamte Kontinuum der Prädiktoren miteinander zu vergleichen. Das Vorgehen wurde ausführlicher in Abschnitt 6.3.4 erläutert. Damit lässt sich veranschaulichen, ob ein bestimmter Prädiktor über das gesamte Kontinuum überlegen ist (globale Überlegenheit eines Prädiktors) oder nur in bestimmten Bereichen (lokale Überlegenheit) besser diskriminiert. Zudem sind ROC-Kurven und insbesondere deren Flächenmaß als Validitätsmaß weniger beeinflusst durch *restriction of range* und Besonderheiten der *Scoreverteilung* als andere Methoden (Humphreys & Swets, 1991).

An dieser Stelle werden wiederum die beiden Interessentengruppen, die entweder ausschließlich Informatik, Elektrotechnik beziehungsweise Technische Informatik ($n = 1582$) oder eine Geisteswissenschaft ($n = 298$) studieren möchten, anhand der Verfahren, aber auch der Abiturdurchschnittsnote eingeteilt. Die Sensitivität und das Komplement der Spezifität (zur Erläuterung siehe Abschnitt 6.3.4) bei jedem möglichen Grenzwert bilden dann die Grundlage für die Abbildung 6.8.

Die nicht-kognitiven Fragebogen weisen dabei nahezu über das gesamte Spektrum möglicher Cut-Offs eine höhere Trennschärfe als die kognitiven Tests auf. Die jeweiligen Flächen unter den *ROC-Kurven* (*area under curve* oder *AUC*) verdeutlichen die Unterschiede:

⁵ Die Berechnung ist anhand des *proportional chance criterion* nach Morrison (1969) erfolgt.

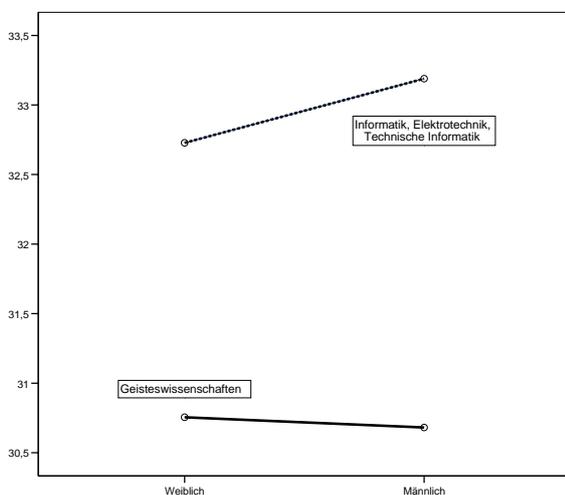


Abbildung 6.1: Geschlechtsunterschiede SW

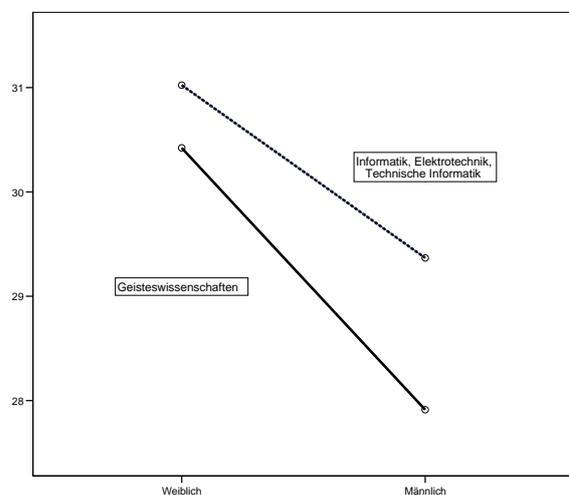


Abbildung 6.2: Geschlechtsunterschiede ELM

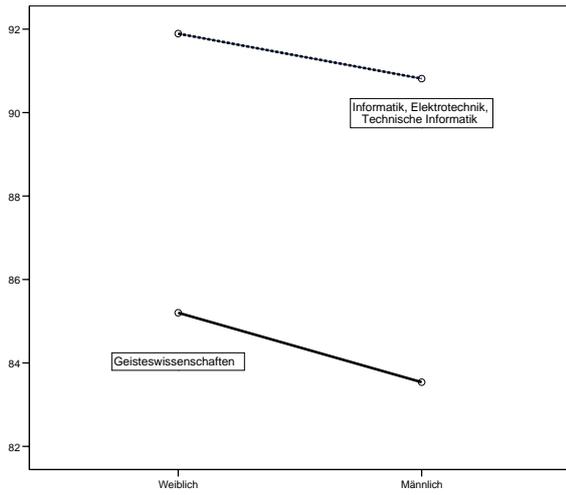


Abbildung 6.3: Geschlechtsunterschiede HK

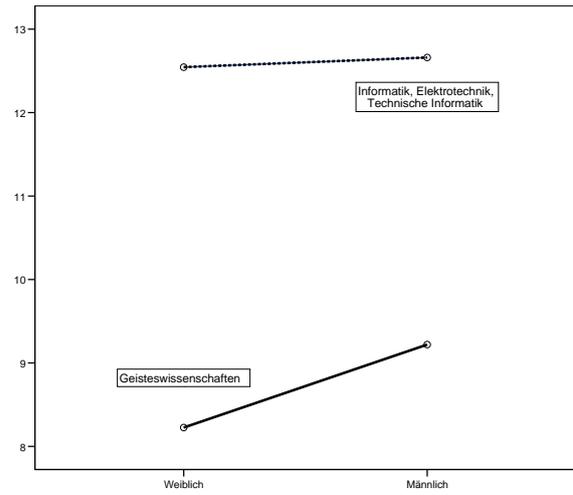


Abbildung 6.4: Geschlechtsunterschiede INT

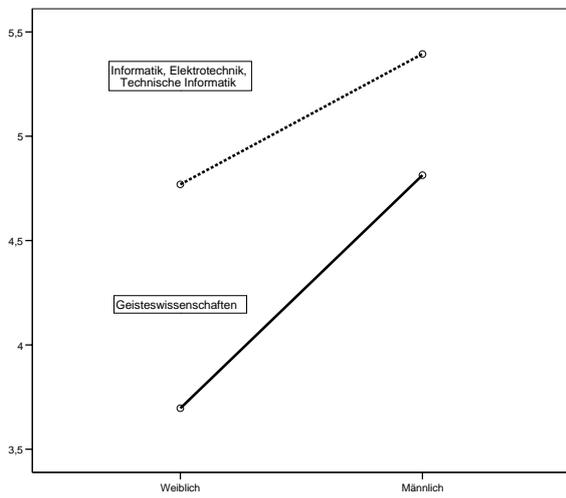


Abbildung 6.5: Geschlechtsunterschiede MAT

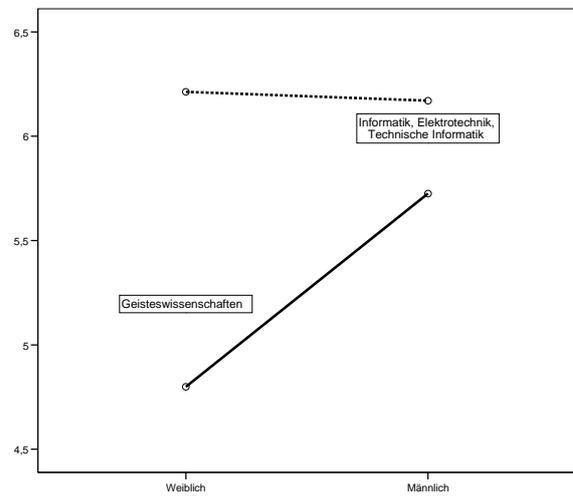


Abbildung 6.6: Geschlechtsunterschiede LOG

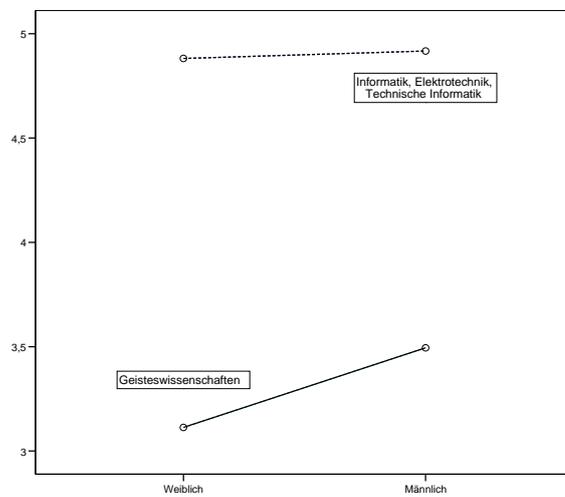


Abbildung 6.7: Geschlechtsunterschiede MATH

Tabelle 6.4
 Logistische Regression zur Vorhersage des Studieninteresses

Verfahren	Nagelkerkes R^2	$-2LL$	χ^2 (Modell)	df	Signifikanz (p)
<i>C-Index</i>	0,174	1443,245	200,597	1	< 0,001
+ <i>HOM, HOP, HOT, ELM, SW</i>	0,264	1330,291	313,550	6	< 0,001
+ <i>LOG, MATH, MAT</i>	0,331	1240,654	403,187	9	< 0,001

Anmerkungen: $n = 1880$

Die nicht-kognitiven Fragebogen ($AUC = 0,798$, 95 %-Konfidenzintervall zwischen 0,773 und 0,824) verursachen weniger Zuordnungsfehler als die kognitiven Tests ($AUC = 0,732$, 95 %-Konfidenzintervall zwischen 0,703 und 0,761). Die gesamte Zuordnungsleistung der Testbatterie mit $AUC = 0,841$ (95 %-Konfidenzintervall zwischen 0,817 und 0,864) liegt weit über dem Zufall (dort wäre $AUC = 0,500$). Die Abiturdurchschnittsnote ist hingegen nicht geeignet, zwischen den Interessentengruppen zu diskriminieren ($AUC = 0,506$, 95 %-Konfidenzintervall zwischen 0,470 und 0,543). Alle gebildeten Flächen sind – bis auf die der Abiturdurchschnittsnote – in ihrer Größe signifikant von $AUC = 0,500$ verschieden.

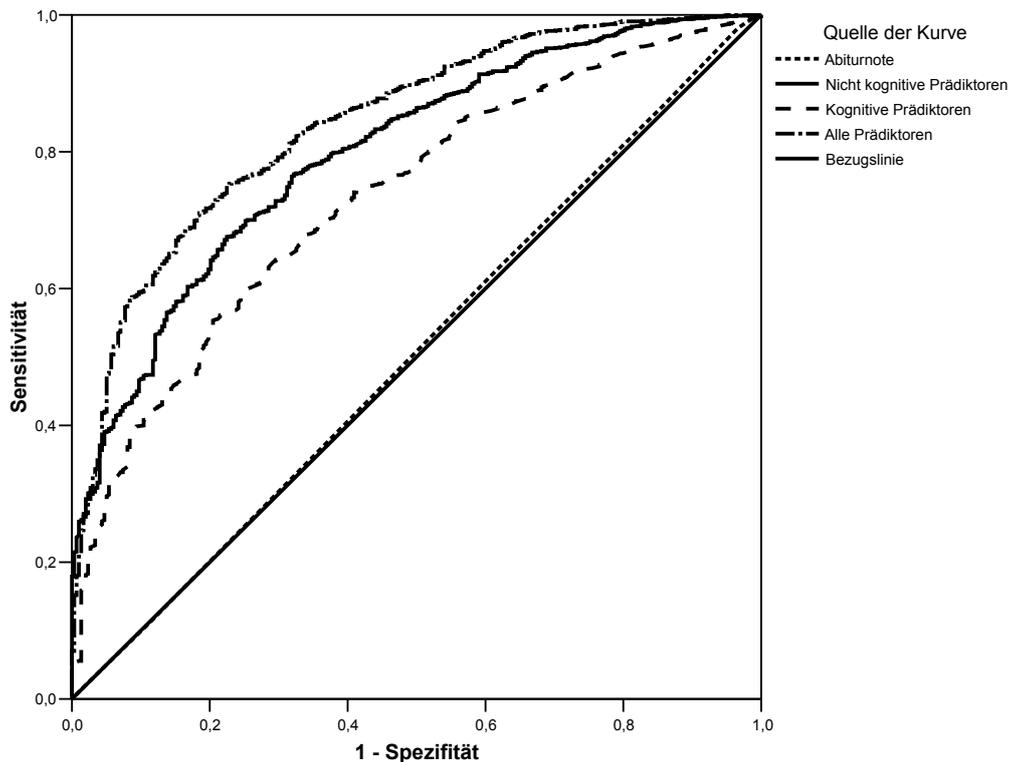


Abbildung 6.8: Sensitivität und Komplement der Spezifität zur Unterscheidung von Interessentengruppen (Studienfachinteresse Geisteswissenschaft versus Informatik/Elektrotechnik)

Betrachtet man die gesamten Testergebnisse als Profilverläufe, so lässt sich noch analysieren, ob signifikante Gestalt- und/oder Höhenunterschiede der Profilverläufe der beiden Präferenzgruppen im Self-Assessment vorhanden sind. Huber (1973) entwickelte auf Basis des Cattell-

schen Profilähnlichkeitskoeffizienten r_p verbesserte Koeffizienten, die zufallskritisch abgesichert und gleichsam auch bei korrelierten Profilskalen eingesetzt werden können. Neben einem globalen Index unterscheidet er zwei Bewertungskriterien: die Höhen- und die Gestaltunterschiede. Beide Koeffizienten können theoretisch zwischen $+1$ (identische Testprofile) und -1 (maximal unterschiedliche Testprofile) schwanken. Hier eingesetzt sollen sie die Unterschiede zwischen den gemittelten Profilen der beiden betrachteten Gruppen quantifizieren: Dabei deuten beide Indizes – also sowohl der für die Höhe als auch der für die allgemeine Gestalt – auf einen sehr hohen Unterschied hin ($g(h) = -1,000$ und $g(v) = -1,000$). Zufallskritisch lässt sich dieses Ergebnis bei $\alpha = 0,050$ absichern. Somit lässt sich zeigen, dass sich die Profile der unterschiedlichen Interessenten signifikant unterscheiden – und zwar im Verlauf und in der Höhe gleichermaßen.

In den bisherigen Analysen wurden die Studieninteressierten der Informatik und Elektrotechnik stets als eine Stichprobe betrachtet, wurden doch im Rahmen der Anforderungsanalysen (siehe Zimmerhofer, 2003) für beide Studienfächer ähnliche Anforderungen definiert und entsprechend dieselben Verfahren verwandt. Im Folgenden soll genauer analysiert werden, ob dennoch Unterschiede in den mittleren Testscores dieser beiden Interessentengruppen bestehen und – wenn ja – in welchen Komponenten der Testbatterie. Betrachtet man die eigentliche Zielgruppe und ihre Scoremittelwerte genauer und differenziert zwischen den Interessierten der Elektrotechnik und der Informatik, so fallen die insgesamt geringen Effektstärken auf (siehe Tabelle 6.1 im Vergleich zu 6.2). Nahezu alle Mittelwertsunterschiede zeigen Werte von unter $d = 0,200$. Nach Cohen (1992) entspricht $d = 0,200$ einer noch geringen Effektstärke. Effektstärken unter dieser Grenze können entsprechend vernachlässigt werden. Allein die Effektstärke für *SW* deutet auf einen größeren Unterschied hin ($d = 0,426$). Aber auch dieser Wert entspricht nach Cohens Klassifikation lediglich einer geringen Effektstärke.

Die geringste Effektstärke ($d = 0,020$) tritt beim Vergleich der mittleren Interessenkongruenzen der Gruppen Informatik und Elektrotechnik auf. Die theoretische Anzahl unterschiedlicher Three-Letter-Codes liegt bei 120. Davon zeigen sich in der Gruppe der an Informatik Interessierten 86, in der Gruppe der an Elektrotechnik Interessierten 77 Codes. Bei beiden Gruppen fallen zwei Three-Letter-Codes auf, die zusammen etwa ein Fünftel beziehungsweise ein Viertel der Personen beschreiben: 11,2 % der Informatik-Interessierten und 14,8 % der Elektrotechnik-Interessierten zeigten das Interessenprofil *I-R-E*. Sie sind also an *wissenschaftlich-analytischen*, dann an *praktisch-technischen* und zuletzt an *unternehmerischen* Tätigkeiten interessiert. Zweithäufigster Three-Letter-Code in beiden Gruppen ist *I-E-R* (8,4 % bei Informatik, 10,3 % bei Elektrotechnik). Betrachtet man nicht die intraindividuell präferierten Dimensionen, sondern die mittleren Skalenscores des Interessenfragebogens, so zeigt sich ebenfalls die besondere Bedeutung dieser drei Dimensionen: Bei beiden Gruppen zeigt sich auf Ebene der Mittelwerte der Three-Letter-Code ebenfalls *I-R-E*. Für die Informatik lassen sich folgende Mittelwerte berechnen: $M = 36,814$ für *INT-I*, $M = 30,374$ für *INT-R*, $M = 26,805$

für *INT-E*. Für die Elektrotechnik gelten vergleichbare Ergebnisse: $M = 35,926$ für *INT-I*, $M = 31,111$ für *INT-R*, $M = 30,757$ für *INT-E*.

Im Rahmen einer erneuten logistischen Regression zur Vorhersage der Gruppenzugehörigkeit *Informatik* oder *Elektrotechnik* sind die bereits beschriebenen Prädiktoren des Self-Assessments lediglich in der Lage, 2,9 % der Varianz (Nagelkerkes R^2) aufzuklären. Das Flächenmaß unter der *ROC*-Kurve (siehe Abbildung 6.9) für alle Prädiktoren beträgt $AUC = 0,580$ (95 %-Konfidenzintervall zwischen 0,508 und 0,577). Wieder ist die Fläche unter der *ROC*-Kurve der nicht-kognitiven Prädiktoren mit $AUC = 0,560$ am größten. Diese Fläche ist aufgrund der hohen Stichprobengröße – ebenfalls wie die Fläche der kognitiven Prädiktoren bei $\alpha = 0,050$, nicht aber die der Abiturdurchschnittsnote – signifikant von $AUC = 0,500$ verschieden.

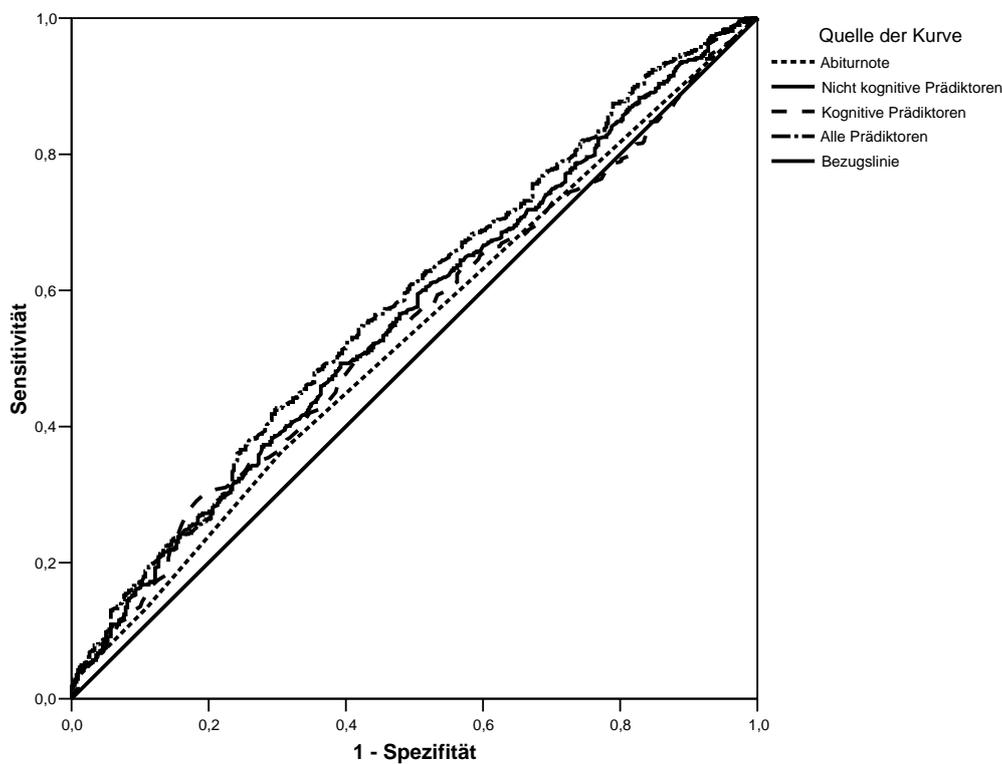


Abbildung 6.9: Sensitivität und Komplement der Spezifität zur Unterscheidung von Interessentengruppen (Studienfachinteresse Informatik versus Elektrotechnik)

6.5 Diskussion

In der zweiten Studie dieser Arbeit sollte überprüft werden, inwieweit die Testbatterie des Self-Assessments in der Lage ist, die eigentliche Zielgruppe des Beratungssystems von einer geisteswissenschaftlich interessierten Gruppe zu unterscheiden.

Im Ergebnisteil wurden erwartungsmäßig bedeutsame Mittelwertsunterschiede zwischen diesen Interessentengruppen aufgezeigt und eine Zuweisung post-hoc aufgrund der Testergebnisse durchführt. Dieses Ergebnis deutet auf folgende Sachverhalte hin:

1. Die beiden Personengruppen unterscheiden sich hinsichtlich ihrer Fähigkeiten und Eigenschaften.
2. Ferner lässt sich ableiten, dass Personen, die sich für ein bestimmtes Studienfach interessieren, auch im Mittel Fächer wählen, die ihren Stärken und Schwächen entsprechen (Selbstselektion).
3. In der Anforderungsanalyse wurden Fähigkeits- und Eigenschaftsbereiche aufgezeigt und im Self-Assessment umgesetzt, die (zumindest bezüglich Informatik und Elektrotechnik) fachspezifisch sind und nicht ausschließlich nur die allgemeine Studierfähigkeit beschreiben.
4. Zudem lässt sich aufzeigen, welche Stärken und Schwächen vom Suchenden ausgewertet und wie diese gewichtet werden, um zu einer individuellen Studienpräferenz zu finden. Diese lässt sich dann – wie hier erfolgt – als artikuliertes Studieninteresse erfragen.
5. Dies deutet ferner darauf hin, dass das artikuliertes Studieninteresse und die individuelle Eignung und Neigung beim Großteil der Studieninteressenten schon deckungsgleich sind. Aber Diskrepanzen zwischen diesen beiden Parametern deuten auf ein unklares Selbstbild oder auf unzureichende Information über Studieninhalte hin (Bergmann, 1994).

Die nicht-kognitiven Verfahren zeigen sich dieser Studie zufolge als die besten Prädiktoren für die Studienpräferenz. Gerade der Interessenfragebogen kann als Sammlung globaler Tätigkeitsbeschreibungen (*measured interests*, siehe Abschnitt 3.4.3) die artikuliertes Studienfachwahl (*expressed interests*) gut vorhersagen. Dies lässt vermuten, dass die individuelle Studienfachwahl – wie bereits postuliert – primär anhand der subjektiv empfundenen Interessenstruktur vorgenommen wird und diese zugleich hoch mit der im Self-Assessment objektiv messbaren Struktur übereinstimmt. Genauso scheinen Interessenten subjektiv aber auch die Eignung zum Beispiel im Bereich der Mathematik berücksichtigen, um die Studienfachwahl zu vollziehen.

Ein weiteres wichtiges Ergebnis ist, dass alle Test- und Fragebogenverfahren in dieser Stichprobe besser in der Lage sind, Unterschiede zwischen den beiden Interessentengruppen aufzuzeigen als die Abiturdurchschnittsnote allein. Wahrscheinlich würde allerdings eine Berücksichtigung der Einzelschulnoten mehr Varianz aufklären als die hier verwendete Abiturdurchschnittsnote. Immerhin wird das Wissen über eigene Stärken und Schwächen primär von der Kenntnis der einzelnen Noten stammen. Auch würde der Abiturdurchschnittsnote eine Trennung besser gelingen, wenn Studienfächer mit unterschiedlichen Zulassungsvoraussetzungen (zum Beispiel einem hohen zentralen Numerus clausus) betrachtet werden würden. Mehr Forschung ist in diesem Bereich nötig, um die Variablen für eine individuelle Studienfachwahl genauer ergründen

zu können. Die hier erläuterte Studie zeigt auf, wie der Verlauf einer Studienfachwahl erkundet werden kann, ohne die Studieninteressenten nach genau dieser befragen zu müssen.

Wie vermutet, zeigen sich im direkten Vergleich der Interessierten der Informatik und der Elektrotechnik keine bis geringe Unterschiede in den Testscores. Diese Studieninteressenten haben dementsprechend also im Mittel vergleichbare Eignungen und Neigungen (zumindest hinsichtlich der hier erhobenen Bereiche); so entstehen vergleichbare Umwelten. Eine getrennte Normierung und Berechnung der Interessen-Kongruenz ist folglich nicht notwendig. Ob eine Auftrennung der beiden Fächer sinnvoll ist, kann an dieser Stelle nicht entschieden werden, fehlt doch noch eine Untersuchung der Kriteriumsvalidität in der Elektrotechnik. So könnte es sein, dass es sehr wohl Unterschiede in der Eignung für die beiden Fächer gibt, diese aber nicht von dem Self-Assessment abgebildet werden können. Sollte geplant sein, getrennte Self-Assessments zur Verfügung zu stellen, um zum Beispiel eine differenziertere Studienberatung zu ermöglichen, so müssten auf jeden Fall weitere Messbereiche nach einer erneuten Anforderungsanalyse aufgenommen werden. Bis auf den Fragebogen der Selbstwirksamkeit gelingt nämlich für keine Skala eine bedeutsame Differenzierung der beiden Interessentengruppen. Mit einer Mittelwertdifferenz von einer halben Standardabweichung ($d = 0,426$) beschreiben sich die Informatik-Interessierten als selbstwirksamer als die Elektrotechnik-Interessierten. Weitere Forschung ist nötig, um die Gründe dafür aufdecken zu können.

Folgende Beschränkungen der Studie müssen noch diskutiert werden:

Im Rahmen der hier berichteten Studie wurden ausschließlich Teilnehmergruppen miteinander verglichen, die Fächer mit sehr unterschiedlichen Studieninhalten (Informatik, Elektrotechnik versus Geisteswissenschaften) präferieren. Interessant wäre daher ein Vergleich mehrerer sowohl ähnlicher als auch unähnlicher Studienfächer. Der Einsatz von Self-Assessments scheint gerade dann von Bedeutung zu sein, wenn Personen sich zum Beispiel nicht zwischen den Fächern Wirtschaftsingenieurwesen, Informatik, Elektrotechnik, Maschinenbau etc., also zwischen verwandten ingenieurwissenschaftlichen Fächern, entscheiden können. Eine solche Profilanalyse hilft, Self-Assessments für weitere Studienfächer zu entwickeln, da die empirischen Ergebnisse aufzeigen, welche Konstrukte trennscharf zwischen Studieninteressierten unterscheiden lassen.

Da das Beratungssystem im Internet frei zugänglich war und gerade die Interessierten insbesondere der Geisteswissenschaften nicht während einer Immatrikulation etc. zur Teilnahme verpflichtet waren, lässt sich nicht klären, inwieweit die Stichprobe repräsentativ für alle Geisteswissenschaftler ist. Anzeichen für eine bestimmte Selbstselektion sind jedoch nicht zu erkennen. Anzunehmen ist, dass eher IT-affine Studieninteressierte der Geisteswissenschaften teilgenommen haben. Dadurch hätten sich aber die Unterschiede (und damit auch die Effektstärken) zu der eigentlichen IT-Zielgruppe eher verkleinern als vergrößern müssen.

Kapitel 7

Studie 3: Kriteriumsvalidierung der Testbatterie

7.1 Einleitung

In der Studie 2 (siehe Abschnitt 6) konnte bereits nachgewiesen werden, dass die Self-Assessment-Testbatterie in der Lage ist, die eigentliche Zielgruppe (Studieninteressierte der Informatik, Elektrotechnik und Technischen Informatik) von Personen zu unterscheiden, die ein geisteswissenschaftliches Studium aufnehmen möchten. Dieses Ergebnis deutet darauf hin, dass die Fragebogen und Tests des Self-Assessments auch fachspezifische und nicht ausschließlich allgemeine Fähigkeiten und Neigungen abbilden. Ebenso scheint – bei aller Vorsicht in der Interpretation –, die Selbstselektion der Studieninteressierten insoweit zu funktionieren, dass sie ihre eigenen Stärken und Schwächen weit mehr folgen und ihre Präferenzen anpassen, als landläufig angenommen wird.

Als zentrales Kriterium für die bisherigen Analysen galt somit stets die Präferenzaussage der Studieninteressenten (also Aussagen zum „Interesse an einem Studium der Informatik“), die zeitgleich prognostiziert werden sollte. Gerade für die Praxis der Studienberatung interessanter erscheint hingegen die Fragestellung, ob das Testergebnis im Self-Assessment vor der Studienaufnahme Vorhersagen über den Studienverlauf zulässt. Anzunehmen ist, dass es einer Testbatterie noch vergleichsweise leicht gelingt, Extremgruppen hinsichtlich ihrer Studienwahl wie berichtet konkurrenzlos zu trennen. Schwieriger erscheint es, Gruppen zu unterscheiden, die homogener in ihrer eigenen Präferenz sind (sich also zum Beispiel alle für Informatik interessieren), sich aber in ihrer Eignung unterscheiden; jedoch genau dort liegt der eigentliche Auftrag eines Beratungssystems. Es soll schließlich helfen, Personen, die Informatik und Elektrotechnik studieren möchten, über ihre Stärken und Schwächen eben genau für dieses Studium aufzuklären.

Die Frage, die in dieser Studie beantwortet werden muss, ist demnach, ob es einem webbasierten System von ca. 90 Minuten Testdauer gelingen kann, eine Vorhersage des Studienerfolgs zu leisten. Operationalisiert wird der Studienerfolg hier sowohl über die mittleren Vordiplomnoten als auch über die Anzahl bestandener Vordiplomprüfungen (alle Noten erfasst bis zum Mai 2006). Somit werden sowohl die Qualität als auch die Quantität von Prüfungsleistungen erfragt. Gerade das letztere Kriterium (siehe auch Abschnitt 3.1) hängt eng mit dem Verhaltensmuster der Procrastination (Prüfungsaufschieben) zusammen (siehe Abschnitt 3.4). Neben dem primären Kriterium eines guten Studienabschlusses wird immer stärker auch die Studiendauer beachtet – und zwar gerade aufgrund der Studiengebühren auch von den Studenten selbst.

Von Interesse sein wird weiter, welche Tests beziehungsweise Fragebogen besonders geeignet sind, den Studienerfolg vorherzusagen. Sollte sich zeigen, dass bestimmte Tests oder Fragebogen keine oder keine zusätzliche Varianz im Kriterium aufklären, müssen inhaltliche Verbesserungen angedacht werden. Auch eine Verkürzung des Self-Assessments um diesen Prädiktor stellt – gerade aufgrund der für Internetverfahren langen Testdauer – einen gangbaren Weg dar.

Sollten bestimmte Prädiktoren unterschiedlich viel Varianz beim Vergleich der beiden Kriterien aufklären, so erscheint dies, ein interessanter Anknüpfungspunkt für weitere Forschungsbemühungen zu sein. Ähnliches gilt für den Fall, dass die Prädiktion der Studienleistungen im Hauptfach besser gelingt als für das Nebenfach der Informatik oder vice-versa.

7.2 Fragestellung

Entsprechend der US-amerikanischen (Burton & Ramist, 2001; Kuncel et al., 2005; Kuncel, Hezlett & Ones, 2001; Robbins et al., 2004) und auch der deutschsprachigen Literatur (Baron-Boldt, Schuler & Funke, 1988; Hell, Trapmann, Weigand, Hirn & Schuler, 2005; Trapmann, Hell, Weigand & Schuler, 2007; Trost, Klieme & Nauels, 1997) ist anzunehmen, dass sowohl die Abiturdurchschnittsnote als auch die kognitiven Tests gute Prädiktoren für die Vordiplomnoten in der Informatik darstellen. Die nicht-kognitiven Fragebogen werden erfahrungsgemäß nur eine geringe, weitere Varianzaufklärung über diese beiden Prädiktoren hinaus bieten können (Bergmann & Eder, 1991; Hell, Trapmann, Weigand, Hirn & Schuler, 2005; Tsabari, Tziner & Meir, 2005).

Auch für die vorhersagte Anzahl bestandener Vordiplomprüfungen wird angenommen, dass alle drei Prädiktorgruppen entscheidend zur Verbesserung der Prädiktion beitragen. Allerdings wird hier vermutet, dass die Varianzaufklärung der nicht-kognitiven Fragebogen (insbesondere der Skalen der Handlungskontrolle *HK*) größer ist als bei der Prädiktion der Vordiplomnoten.

Dabei ist die inhaltliche Nähe der Procrastination (Helmke & Schrader, 2000, siehe auch Abschnitt 3.4.6) zum Kriterium entscheidend. Trotzdem wird die Varianzaufklärung dieser Skalen geringer ausfallen als die der kognitiven Skalen und der Abiturdurchschnittsnote.

7.3 Methoden

In der Studie 3 wird – anders als Studie 1 – eine spezifische Personengruppe behandelt, bei der es möglich war, angefallene Prüfungsnoten für eine Kriteriumsvalidierung zu erheben. Dabei wurden die Noten über einen Anonymisierungsserver pseudonymisiert gemeldet. Die Studenten mussten die Noten also nicht mittels einer Selbstauskunft melden.

7.3.1 Stichprobe

Zu Beginn des Wintersemesters 2004/2005 war die Teilnahme am kompletten Self-Assessment verpflichtend für alle Studierende der Vorlesungen „Programmierung“ und „Programmierung für Alle“ an der RWTH Aachen. Hauptzielgruppe der Vorlesung „Programmierung“ (Umfang: vier Semesterwochenstunden) waren Studierende der Diplom-Informatik (Hauptfach) im ersten Semester. Die Vorlesung „Programmierung für Alle“ (Umfang: zwei Semesterwochenstunden) wird für Nebenfachstudierende der Informatik gehalten. Die beiden Vorlesungen wurden ausgewählt, da sie zu Beginn des Studiums absolviert werden müssen. Sie stellen den Erstkontakt zur Hochschule dar. Die Testteilnahme konnte zu Hause erfolgen, genauso aber auch im Computerpool der Informatik. Technischer Support wurde soweit erforderlich direkt per Mail durch den Autor dieser Arbeit geleistet.

Im Hauptfach stehen für 182 Studierende Kriteriumsdaten, die elektronisch und pseudonymisiert direkt von den Lehrstühlen gemeldet wurden, zur Verfügung. Im Mittel war diese Stichprobe 21,2 Jahre alt ($SD = 2,423$). 10 % der Teilnehmer waren weiblich. Im Nebenfach liegt die Stichprobengröße bei $n = 412$ mit einem Altersmittel von 20,9 Jahren ($SD = 2,102$). 17 % dieser Studierenden sind weiblich. 70 % dieser Nebenfachstudenten studieren im Hauptfach Wirtschaftsingenieurwesen, 14 % Physik, 10 % Technik-Kommunikation, 2 % Operations Research und 2 % Biologie. Alle Prüfungsleistungen, die bis Mai 2006 dem Prüfungsamt der RWTH Aachen übermittelt wurden, sind in diese Analyse eingeflossen.

Das Data-Screening wurde analog zu den Beschreibungen in Abschnitt 5.3.1 vorgenommen.

7.3.2 Verwandte Testverfahren

Die Verfahren entsprechen den bereits in Studie 1 (siehe Abschnitt 5.3.2) beschriebenen Skalen. Alle Testungen wurden mit der vierten, also letzten Testrevision durchgeführt.

7.3.3 Technische Realisierung

Ein wesentliches technisches Instrument für die Durchführung der Studie 3 war ein so genannter Anonymisierungsserver, der im Jahr 2004 vom Autor dieser Arbeit und Mitarbeitern der Informatik der RWTH Aachen (insbesondere aus der Computer-Supported Learning Research Group) entwickelt und 2005 überarbeitet wurde. Der Anonymisierungsserver ermöglichte, besonders sensible, personenbezogene Informationen auf zwei getrennten Webservern zu speichern und dennoch eine Kriteriumsvalidierung zu ermöglichen. Der Self-Assessment-Server konnte so die Testdaten aus dem Self-Assessment verwalten, hatte aber keinen Zugriff auf Matrikelnummern oder Namen. Die einzelnen Server der Informatiklehrstühle, die die Notenverwaltung übernahmen, hatten ihrerseits keine Informationen über die Testteilnahmen (Testdaten, Studienzufriedenheit etc.), hingegen die Namen, die Prüfungsnoten und die Matrikelnummern. Der Anonymisierungsserver vermittelte zwischen diesen beiden Servern und ermöglichte das Zusammenführen von Testlauf und Studiennoten, ohne den Datenschutz zu gefährden. Um ein Zusammenführen der Datensätze und damit eine Kriteriumsvalidierung zu ermöglichen, war in den jeweiligen Datenbanken ein eindeutiger Schlüssel gespeichert. Direkte Rückschlüsse, zum Beispiel auf die Matrikelnummer, waren damit nicht möglich.

Anders als die Studiennoten, die von den Lehrstühlen gemeldet wurden, sind die Abiturnoten vom Teilnehmer erfragt und nicht über das Prüfungsamt etc. eingeholt worden. Auch dabei bestand kein Grund, ein gesteigertes Impression-Management anzunehmen, da diese Schulnoten nicht an die Lehrstühle weitergegeben wurden (und dies auch an mehreren Stellen in den Instruktionen verbalisiert war). Kompletต์ ausschließen lässt sich eine Verzerrung aufgrund der Selbstauskunft aber nicht.

Dickhäuser und Plenter (2005) wiesen für die TIMS-Studie eine hohe Korrelation zwischen den selbst berichteten Schulnoten und den wahren Noten der Lehrer nach. In einer US-amerikanischen Metaanalyse wurden zwar höhere Abweichungen aufgezeigt. Aber auch dort hat die Verzerrung die prognostische Validität der selbst berichteten Note für das spätere Studium nicht bedeutsam gesenkt (Kuncel, Crede & Thomas, 2005).

7.3.4 Verwandte Auswertungsverfahren

Die angestrebte Kriteriumsvalidierung wird auf Basis multipler linearer Regressionen vorgenommen (Backhaus, Erichson, Plinke & Weiber, 2006, Bortz, 1999). Damit sollen Gleichungen optimal gewichteter und linear kombinierter Prädiktoren gefunden werden, die die Vorhersage der mittleren Studiennoten und der Anzahl der bestandenen Prüfungen anhand der Testergebnisse ermöglichen. Ein besonderes Problem bei diesem Vorgehen besteht darin, dass die Konsistenz und somit die Generalisierbarkeit dieser regressionsanalytischen Ergebnisse durch eine optimale Gewichtung auch von Zufallsfehlern in der betrachteten Stichprobe beeinträchtigt werden kann. Folglich können Ergebnisse von Regressionsanalysen – inklusive der β -Gewichte und des Bestimmtheitsmaßes – gerade bei kleineren Stichproben stark differieren. Die Überschätzung der Prädiktoren-Kriterium-Beziehung fällt insbesondere dann ins Gewicht, wenn – wie in diesem Fall – zahlreiche Prädiktoren verwandt werden, da entsprechend mit jedem Prädiktor auch mehr Möglichkeiten bestehen, von zufälligen Beziehungen der Variablen zum Kriterium Gebrauch zu machen (Diehl & Kohr, 1999).

Um diese Gefahr einschränken zu können, werden zwei unterschiedliche Methoden verwandt, eine Kreuzvalidierung durchzuführen: Das klassische Vorgehen besteht darin, den gesamten Datensatz in zwei Hälften zu teilen (Diehl & Kohr, 1999; Diehl & Staufenbiel, 2001). Somit stehen für die Regressionsanalyse im Hauptfach $n = 91$ beziehungsweise $n = 206$ im Nebenfach zur Verfügung. Trotz dieser Aufteilung liegt die Stichprobengröße in beiden Fällen immer noch über der mittleren Größe von publizierten Kriteriumsvalidierungen (Schmidt, 1992). Die berechnete Regressionsgerade, die auf der Basis der ersten Stichprobe ermittelt wurde, wird für die zweite Hälfte des Datensatzes zur Prädiktion verwandt. Eine Korrelation zwischen dem Kriterium und der Vorhersage auf Basis dieser Regressionsgeraden lässt somit die Generalisierbarkeit der Gewichte abschätzen. Die zentrale Problematik der starken Stichprobenabhängigkeit wird durch diese klassische Vorgehensweise aber nicht schlussendlich gelöst, stellen die beiden zufälligen Aufteilungen nur eine spezifische Auswahl vieler möglicher Stichprobenkonstellationen dar. Zu beachten ist ferner, dass die Stichprobe für diese Kreuzvalidierung halbiert, also zusätzlich verkleinert wird, so dass Zufallsfehler gerade der β -Gewichte eher bedeutsamer ausfallen.

Zusätzlich zu diesem klassischen Vorgehen wird mit dem Bootstrapping eine zweite Technik zur Kreuzvalidierung verwandt. Zur allgemeinen Einführung in das Thema wird an dieser Stelle auf Efron und Tibshirani (1993) verwiesen. Bei diesem Vorgehen werden aus der gegebenen Stichprobe keine zwei, sondern eine große Anzahl unterschiedlich zusammengesetzter Stichproben mit Zurücklegen zufällig gezogen. Entsprechend werden viele verschiedene Regressionsanalysen berechnet und die Ergebnisse (allen voran die multiple Korrelation, die Bestimmtheitsmaße und die standardisierten Koeffizienten) in Histogrammen dargestellt, so dass sich ihre Verteilung auf Basis der zufällig gezogenen Stichproben abschätzen lässt. Efron and Tibshirani

(1993) empfehlen 500 bis 1 000 Zufallsziehungen vorzunehmen, um die Schätzfehler ausreichend klein zu halten. Entsprechend werden hier jeweils 1 000 Zufallsziehungen vorgenommen. Die Ausgangsbasis der hier verwandten Implementierung bestand aus einem Macro von SPSS Inc. (Levesque, 2005), das zusätzlich vom Autor dieser Arbeit auf die Fragestellung angepasst wurde: So wird vor der Mittelung der Korrelationskoeffizienten bei der modifizierten Version eine Fisher-z-Transformation durchgeführt (Bortz, 1999). Eine Korrektur der möglichen Varianzeinschränkung der Stichprobenverteilungen im Vergleich zur Verteilung in der Population wird nicht durchgeführt.

Um die hier berechneten Korrelationskoeffizienten in ihrer Bedeutung einschätzen zu können, wird auf die Klassifikation von Cohen (1988) und auf die Vergleichstabelle von Meyer et al. (2001) zurückgegriffen. Gemäß Cohen werden Korrelationen um $r = 0,10$ als gering, $r = 0,30$ als mittel und $r = 0,50$ als groß eingeschätzt. Meyer et al. (2001) sammelten zahlreiche Korrelationskoeffizienten aus der Psychologie und Medizin und erlauben so anschauliche Vergleiche der hier diskutierten statistischen Zusammenhänge.

Zusätzlich werden – wie auch in Studie 2 (siehe Abschnitt 6) – *receiver operating characteristic*-Kurven berechnet. Das in diesem Kontext zu entdeckende „Signal“ ist die Eignung für das Studium. Da in dieser Studie keine Studienabbruchdaten vorliegen, wird ein Split der Gruppe hinsichtlich der Vordiplomnote vorgenommen: Personen, die eine mittlere Vordiplomnote von 4,0 oder schlechter aufzeigen, werden zu den eher nicht erfolgreichen Studenten gezählt. Personen, die Vordiplomprüfungen mit einer besseren Note als 4,0¹ bestanden haben, werden als erfolgreich klassifiziert. Als Prädiktoren, die das entsprechende Signal entdecken sollen, werden sowohl die gesamte Testbatterie (kognitive Testverfahren und nicht-kognitive Fragebogen) als auch die Abiturdurchschnittsnote verwandt. Die Nullhypothese (H_0) im Rahmen der Signalentdeckungstheorie postuliert, dass eine Person nicht für das Studium der Informatik geeignet ist. Die Arbeitshypothese (H_1) geht hingegen davon aus, dass eine Person geeignet ist und somit auch eine bessere Vordiplomnote als im Mittel 4,0 erreicht. Um die jeweiligen Prädiktorenscores zu bilden, wurde auf die oben berichtete Regressiongerade zurückgegriffen.

Entsprechend der Signalentdeckungstheorie werden folgende vier Gruppen der Entscheidung betrachtet:

- Personen werden korrekterweise als nicht geeignet eingeschätzt (richtig-negativ)
- Personen werden fälschlicherweise als nicht geeignet eingestuft (falsch-negativ)
- Personen werden korrekterweise als geeignet eingeschätzt (richtig-positiv)
- Personen werden fälschlicherweise als geeignet eingeschätzt (falsch-positiv)

Die Sensitivität ist in diesem Fall definiert als Anteil der korrekt erkannt Geeigneten (*richtig-positiv*) an der Gesamtzahl der erkannt (*richtig-positiv*) oder unerkannt Geeigneten (*falsch-*

¹ Diese Aufteilung entspricht gleichzeitig einem Mediansplit.

negativ) (siehe Formel 6.1 auf Seite 116). Die Spezifität entspricht hier dem Anteil der korrekt erkannt Ungeeigneten (*richtig-negativ*) unter den erkannt (*richtig-negativ*) oder nicht erkannt Ungeeigneten (*falsch-positiv*) (siehe Formel 6.2). Bei der ROC-Kurve wird die Sensitivität gegen den Komplementärwert der Spezifität zu 1 aufgetragen. Dies entspricht hier den erkannt Geeigneten (*falsch-positiv*) unter den erkannt (*richtig-negativ*) oder unerkannt Ungeeigneten (*falsch-positiv*) (siehe Formel 6.3).

7.4 Ergebnisse

Die dritte hier berichtete Studie soll insbesondere die Einsatzmöglichkeiten der Self-Assessment-Batterie zur Vorhersage des Studienerfolgs prüfen. Diese Analysen basieren – wie bereits beschrieben (siehe Abschnitt 7.3) – auf einer Vollerhebung unter Studenten im ersten Fachsemester der Informatik aus dem Wintersemester 2004/2005. Dabei werden im Folgenden die Hauptfach- und Nebenfach-Studenten getrennt analysiert. Im Mittelpunkt des Interesses stehen hier als Studienerfolgskriterien sowohl die mittleren Studiennoten im Vordiplom als auch die Anzahl der bestandenen Prüfungen. Dazu wurden alle im Prüfungsamt erfassten Prüfungsleistungen bis Mai 2006 ausgewertet.

7.4.1 Korrelationen der Kriterien mit den Prädiktoren

Zu Beginn werden Pearson-Korrelationen zwischen den Studienerfolgskriterien und den Prädiktoren, also den Tests und Fragebogen des Self-Assessments, aber auch der Abiturdurchschnittsnote, der Englischnote, der Informatiknote und der Physiknote der Schule berichtet (siehe Tabelle 7.1). Zur Interpretation der statistischen Beziehungen ist zu beachten, dass die Schulnoten, aber auch die Studiennoten gegenläufig zu den Test- und Fragebogenscores und der Anzahl der bestandenen Prüfungen skaliert sind. Der Logiktest (*LOG*) zeigt sich hier mit $r = -0,404$ ($n = 182$) als bester Einzelprädiktor für die späteren Studiennoten im Hauptfach. Kurzum: Je höher der erreichte Testwert im Logiktest vor dem Beginn des Studiums war, desto wahrscheinlicher zeigte sich später eine gute (also numerisch niedrige) mittlere Vordiplomnote. Dieser Zusammenhang zwischen Test und Vordiplomnote ist sogar enger als der zwischen der Abiturdurchschnittsnote und der Vordiplomnote ($r = 0,386$, $n = 182$). Allerdings gilt letzteres Ergebnis nur in der Hauptfach-Stichprobe: Bei den Nebenfach-Studenten zeigt sich die Abiturdurchschnittsnote im Vergleich zu allen Test- und Fragebogenskalen als besserer Einzelprädiktor. Dort ist außerdem der Mathetest ($r = -0,379$) leicht besser geeignet für die Vorhersage der Studiennoten als der Logiktest ($r = -0,249$).

In beiden Stichproben gilt gleichermaßen: Im Vergleich der Einzel-Schulnoten untereinander gelingt die Vorhersage mit der Mathematiknote ($r = 0,365$ bei den Hauptfachstudenten bzw.

$r = 0,397$ bei den Nebenfachstudenten) am Besten. Die Informatiknote der Schule ($r = 0,145$ bei den Hauptfachstudenten bzw. $r = 0,085$ bei den Nebenfachstudenten) klärt vergleichsweise wenig Varianz im Kriterium auf. Scheinbar hat die Informatik an der Schule wenig mit der Informatik an einer Hochschule zu tun. Im Hauptfach ist auch noch die Englischnote ein bedeutsamer Prädiktor ($r = 0,233$), im Nebenfach hingegen weniger ($r = 0,109$).

Unter den nicht-kognitiven Skalen klärt das Interesse in der Stichprobe der Hauptfach-Studenten die meiste Varianz auf ($r = -0,226$), nicht aber im Nebenfach ($r = -0,066$, $n = 412$). Bei den Nebenfach-Studenten korrelieren die extrinsische Lernmotivation ($r = -0,284$) und die Handlungsorientierung bei Tätigkeitsinitiierung *HOT* ($r = -0,211$) bedeutsam mit den Vordiplomnoten. Bei beiden Stichproben gilt aber, dass die kognitiven Tests in der Prädiktion der Studiennoten den Fragebogenskalen klar überlegen sind.

Betrachtet man die Anzahl bestandener Prüfungen als Studienerfolgskriterium, so verändert sich dieser Sachverhalt zugunsten der nicht-kognitiven Fragebogen nur unwesentlich: Weiterhin bleiben die kognitiven Skalen die wichtigsten Prädiktoren für den Studienerfolg. Drei nicht-kognitive Skalen, nämlich *INT* ($r = 0,238$), *HOT* ($r = 0,178$) und *SW* ($r = 0,202$), sagen bedeutsam die Zahl bestandener Prüfungen voraus. Wieder klärt die Abiturdurchschnittsnote ($r = 0,386$) mehr Varianz auf als alle Test- und Fragebogenverfahren. Die Mathematiknote ($r = 0,397$) korreliert unwesentlich stärker mit dem Kriterium als die Gesamtnote.

Die bisher beschriebenen Ergebnisse zeigen auf, dass die singuläre Prädiktion der unterschiedlichen Erfolgskriterien unterschiedlich erfolgreich verläuft. Gleichzeitig zeigen sich bedeutsame andere Korrelationen zwischen den Prädiktoren und den Kriterien, wenn man die beiden unterschiedlichen Stichproben vergleicht.

Eines gilt für alle Skalen: Sie sind nicht beziehungsweise nur kaum in der Lage, die Anzahl bestandener Prüfungen im Nebenfach bedeutsam vorherzusagen. Die höchste Korrelation mit diesem Kriterium zeigt wieder die Abiturdurchschnittsnote ($r = -0,209$, $n = 412$). Die Mathematiknote ($r = -0,162$) und der Mathetest ($r = 0,190$) zeigen die nächst niedrigeren Korrelationskoeffizienten.

Tabelle 7.1
 Statistischer Zusammenhang (Pearson-Korrelation) zwischen den Prädiktoren (Zeilen) und den Kriterien des Studienerfolgs (Spalten)

	Hauptfach		Nebenfach	
	Vordiplomnoten	Anzahl bestandener Prüfungen	Vordiplomnoten	Anzahl bestandener Prüfungen
Abiturdurchschnittsnote	0,386 ***	-0,386 ***	0,424 ***	-0,209 *
Mathematiknote	0,365 ***	-0,463 ***	0,397 ***	-0,162 *
Englischnote	0,233 **	-0,187 *	0,109	-0,037
Informatiknote	0,145 *	-0,118	0,085	-0,094
Physiknote	0,144	-0,147 *	0,265 *	-0,119
<i>LOG</i>	-0,404 ***	0,373 ***	-0,249 **	0,127
<i>MATH</i>	-0,326 ***	0,276 **	-0,379 ***	0,190 *
<i>MAT</i>	-0,302 ***	0,264 **	-0,280 ***	0,158 *
<i>C-Index</i>	-0,226 **	0,238 *	-0,066	0,125
<i>SW</i>	-0,168 *	0,202 *	-0,136	0,046
<i>ELM</i>	-0,081	0,128 *	-0,284 **	0,094
<i>HOP</i>	0,072	0,000	-0,069	0,004
<i>HOT</i>	-0,049	0,178	-0,211 *	0,133
<i>HOM</i>	0,029	0,041	0,055	-0,056

Anmerkungen: $n = 182$ für Hauptfach-Stichprobe, $n = 412$ für Nebenfach-Stichprobe. Skalenbezeichnungen sind im Abkürzungsverzeichnis erläutert. Auf eine Angabe der Signifikanz wird verzichtet (siehe Abschnitt 5.3.4). Stattdessen werden die Korrelationen in die Effektgröße d transformiert. Ein Asterisk deutet auf eine geringe Effektstärke hin ($0,2 \leq d < 0,5$), zwei Asterisken deuten auf eine mittlere ($0,5 \leq d < 0,8$) und drei auf eine große ($d \geq 0,8$).

7.4.2 Vorhersage der Studiennoten im Hauptfach Informatik

Voraussetzungen der Regressionsanalyse

Vor der Durchführung der Regressionsanalysen zur Prädiktion des Studienerfolgs wird hier exemplarisch für die erste Regressionsanalyse die Überprüfung der Voraussetzungen – unter anderem sollte keine Heteroskedastizität und Multikollinearität nachweisbar sein (Bortz, 1999; Bühner, 2004) – berichtet.

Die Annahme gleicher Residuumsvarianzen bei der multiplen Regression werden sowohl grafisch als auch mithilfe eines Signifikanztests überprüft. Die standardisierten Residuen verteilen sich in der beschriebenen Stichprobe der Hauptfachstudenten Diplom-Informatik erwartungsgemäß annähernd normal (siehe Abbildung 7.1). Bei einer weiteren grafischen Veranschaulichung zur Überprüfung auf Heteroskedastizität wird auf der Ordinate das standardisierte Residuum und auf der Abszisse der standardisierte geschätzte Wert der multiplen Regression abgetragen. In der entsprechenden Abbildung 7.2 lässt sich kein linearer Zusammenhang der Residuen mit den Prognosewerten erkennen. Entsprechend liegt die statistische Beziehung dieser beiden Größen bei 0. Es kann also weiterhin davon ausgegangen werden, dass die Residuumsvarianzen gleich sind.

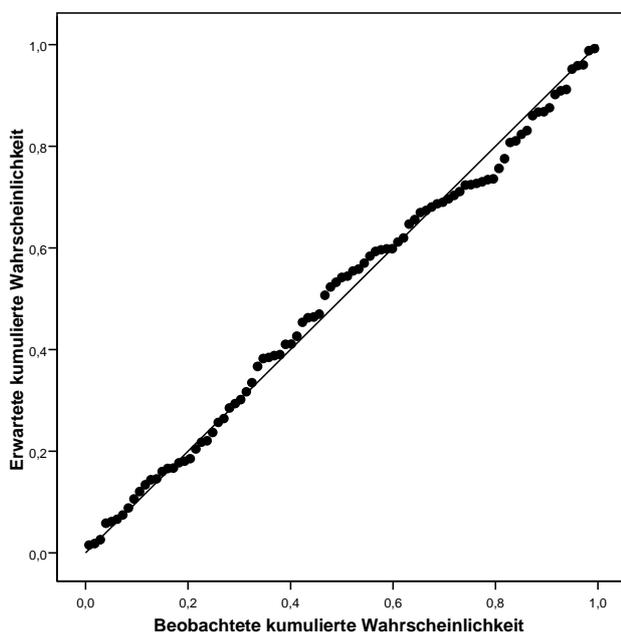


Abbildung 7.1: Prüfung der Residuen auf Normalverteilung

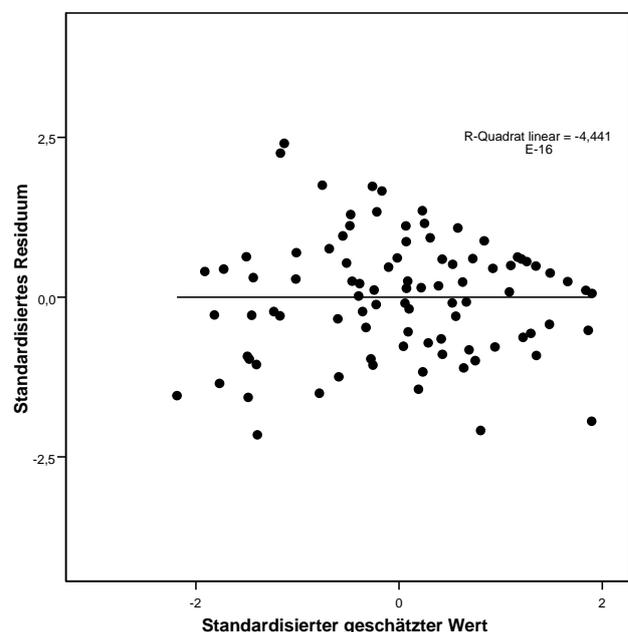


Abbildung 7.2: Prüfung der Residuen auf Heteroskedastizität

Dieser Eindruck wird mithilfe des White-Signifikanztests (White, 1980) gestützt. Die dabei formulierte Nullhypothese, ein Vorliegen der Homoskedastizität, wird in diesem Fall auch bei einer Irrtumswahrscheinlichkeit von $\alpha = 10\%$ nicht verworfen. Somit kann weiterhin von Homoskedastizität ausgegangen werden.

Die Kontrolle auf Multikollinearität, also auf eine lineare Abhängigkeit der Prädiktoren, erfolgt anhand der berechneten Toleranzen und dem Variance Inflation Factor (Backhaus, Erichson, Plinke & Weiber, 2006). Für die Abschätzung der Toleranz wird eine Regression jeder unabhängigen Variable auf die übrigen Variablen mit einer Berechnung des Bestimmtheitsmaßes verbunden. Der dabei berechnete Anteil gemeinsamer Varianz wird dann von 1 subtrahiert. Auch wenn keine exakten Grenzen für eine wenig bedeutsame Multikollinearität in der einschlägigen Literatur existieren, werden Toleranz-Werte oberhalb von 0,9 als ausreichend beschrieben (Backhaus, Erichson, Plinke & Weiber, 2006). Dieser Wert wird hier für alle Variablen überschritten. Auf eine Überprüfung der Autokorrelation (zum Beispiel mit dem Durbin-Watson-Test) wird verzichtet, da keine Zeitreihen eines einzelnen Teilnehmers betrachtet werden.

Da auch für die weiteren Regressionsanalysen keine bedeutsamen Verletzungen der Voraussetzungen gefunden wurden, wird aus Platzgründen auf eine ausführliche Darstellung verzichtet.

Kriteriumsvalidierung mit Kreuzvalidierung

Im Folgenden werden zwei unterschiedliche Formen der Kreuzvalidierung durchgeführt. Für eine klassische Kreuzvalidierung (vgl. Diehl & Staufenbiel, 2001) wird der Datensatz (insgesamt $n = 182$) zufällig in zwei Hälften geteilt (jeweils $n = 91$).

Eine genauere Aussage über die Stichprobenabhängigkeit der multiplen Korrelation und dessen Konfidenzintervall lässt sich mithilfe der Bootstrapping-Methode fällen. Das Vorgehen wurde bereits in Abschnitt 7.3.4 ausführlich beschrieben. Dabei werden hier in dieser Analyse nicht nur zwei, wie in der klassischen Kreuzvalidierung, sondern insgesamt 1 000 zufällige Stichproben mit Zurücklegen aus den 182 Datensätzen gezogen, um die Ergebnisse besser gegen zufällige Schwankungen abzusichern. Alle berücksichtigten Datensätze unterscheiden sich in ihrer Zusammensetzung (d. h. welche Testdaten aufgenommen werden), aber auch in der Gewichtung der einzelnen Daten. Aufgrund des verwandten Algorithmus schwankt die Stichprobengröße zwischen $91 \leq n \leq 124$ (O'Connor, 2000).

Nachfolgend werden zudem im Rahmen der klassischen Kreuzvalidierung in einer hierarchischen multiplen Regression zuerst die kognitiven Tests (*LOG*, *MATH*, *MAT*), danach im zweiten Schritt die nicht-kognitiven Verfahren (*SW*, *ELM*, *HOT*, *HOM*, *HOP*, *C-Index*) zur Prädiktion der mittleren, gleich gewichteten Vordiplomnote im Hauptfach Diplom-Informatik herangezogen. Sowohl die Aufnahme der Prädiktoren im Rahmen des ersten Regressionsmodells als auch die weitere Aufnahme der nicht-kognitiven Verfahren verbessern die Prädiktion bei $\alpha = 0,05$ signifikant (siehe auch Tabelle 7.2). Die multiple Korrelation der kognitiven Subtests mit den gemittelten Vordiplomnoten beträgt $R = 0,445$ ($R^2 = 0,198$). Diese Korrelation verbessert sich

durch die Aufnahme der nicht-kognitiven Verfahren um $\Delta R = 0,117$. Der gemeinsame Varianzanteil von 19,8 % zwischen den kognitiven Tests und den Prüfungsleistungen wird durch die nicht-kognitiven Bestandteile um 59,0 % auf 31,5 % gesteigert.

Wird das Regressionsmodell (siehe Tabelle 7.3) mit den kognitiven und nicht-kognitiven Verfahren als Prädiktoren näher betrachtet, so zeigen sich als signifikante Einzelprädiktoren lediglich der Test zum logischen Schlussfolgern (*LOG*), Fragebogen zur Handlungsorientierung bei der Tätigkeitsausführung (*HOT*) und bei der Handlungsplanung (*HOP*). Wesentlich dabei ist, dass sich acht der neun Prädiktoren mit den Studiennoten gegenläufig verhalten. Höhere Testscores führen entsprechend im Mittel zu besseren Studiennoten (und somit zu einem niedrigeren Notenwert). Lediglich die Skala der Handlungsorientierung bei Planung zeigt ein positives β -Gewicht und somit eine gleichläufige Beziehung mit der Studiennote.

Für die klassische Kreuzvalidierung werden im folgenden Schritt die Koeffizienten, die sich anhand der ersten Teilstichprobe als optimal erwiesen haben, zur Prädiktion in der zweiten Datensatzhälfte verwandt. Bei der Verwendung dieser Regressionsgewichte ergibt die Korrelation aufgrund der Zufallsschwankungen statt $R = 0,562$ nur einen Wert von $R = 0,344$. Die aufgeklärte Varianz beträgt entsprechend 12,0 %.

Die Abbildung 7.3 visualisiert als Ergebnis des Bootstrappings die Verteilung der multiplen Korrelation R zwischen den Prädiktoren und der mittleren Vordiplomnote. Das arithmetische Mittel liegt korrigiert (Bortz, 1999; siehe Fisher-z-Transformation in Abschnitt 7.3.4) bei $R = 0,530$ und $R^2 = 0,280$. Das 95 %-Konfidenzintervall ist für die multiple Korrelation zwischen 0,434 und 0,622, für das Bestimmtheitsmaß zwischen 0,188 und 0,386. Die berichtete multiple Korrelation im Rahmen der klassischen Kreuzvalidierung $R = 0,344$ liegt somit außerhalb des im Bootstrapping berechneten Konfidenzintervalls.

Die kognitiven Prädiktoren zeigen eine multiple Korrelation zu den Vordiplomnoten von $R = 0,464$ (Konfidenzintervall $0,352 \leq R \leq 0,562$), die nicht-kognitiven eine Korrelation von $R = 0,368$ (Konfidenzintervall $0,260 \leq R \leq 0,477$). Die standardisierten β -Gewichte – ebenfalls im Rahmen des Bootstrappings erhoben – können der Tabelle 7.4 entnommen werden. Insbesondere *LOG*, *MATH* und *HOP* zeigen hohe β -Gewichte. Auffällig ist, dass die Handlungsorientierung nach Planung ein positives β -Gewicht zeigt. Somit verschlechtert sich die

Tabelle 7.2
Vorhersage der mittleren Vordiplomnote im Hauptfach, getrennt nach kognitiven und nicht-kognitiven Bestandteilen

Beschreibung	R	R^2	Korr. R^2	F	df	Signifikanz
Kognitive Bestandteile	0,445	0,198	0,171	7,167	3/87	< 0,001
+ nicht-kognitive Bestandteile	0,562	0,315	0,239	4,145	9/81	< 0,001

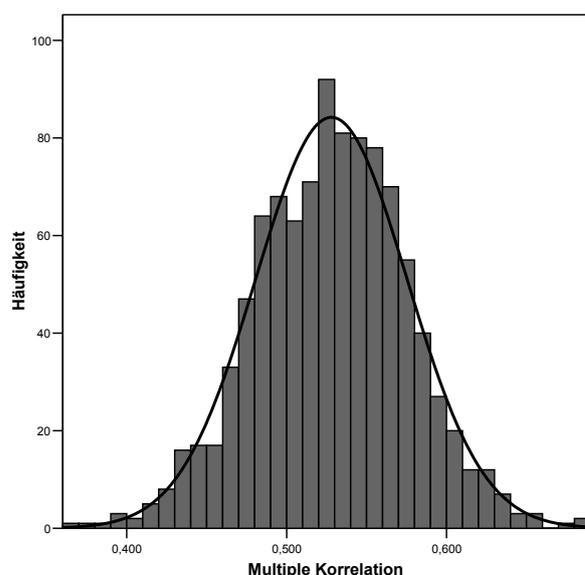
Anmerkungen: $n = 91$; in der Spalte df werden zuerst die Freiheitsgrade für die erklärte, dann für die nicht erklärte Streuung aufgeführt.

Tabelle 7.3

Vorhersage der mittleren Vordiplomnote im Hauptfach Informatik, durch Tests und Fragebogen getrennt

Beschreibung	b	Standardfehler	β -Gewicht	t	Signifikanz
Konstante	8,320	1,166	0,000	7,137	< 0,001
LOG	-0,167	0,058	-0,367	-2,868	0,005
MAT	-0,039	0,049	-0,086	-0,802	0,425
MATH	-0,004	0,066	-0,006	-0,056	0,956
HOP	0,093	0,034	0,538	2,716	0,008
HOT	-0,089	0,031	-0,470	-2,903	0,005
ELM	-0,052	0,028	-0,233	-1,833	0,070
C-Index	-0,023	0,024	-0,095	-0,967	0,336
HOM	-0,019	0,023	-0,087	-0,832	0,408
SW	-0,019	0,035	-0,066	-0,550	0,584

Anmerkungen: n = 91



Beschreibung	R
Mittelwert (korrigiert)	0,530
Konfidenzintervall (95 %)	0,434 0,622

Abbildung 7.3: Verteilung der multiplen Korrelation zwischen der Testbatterie und der mittleren Studiennote. Ermittelt mithilfe von 1000 Bootstrapping-Datensätzen. Die Korrelationen sind vor der Mittelung Fisher-z-transformiert worden.

Tabelle 7.4

Standardisierte β -Gewichte zur Vorhersage der Studiennoten (Bootstrapping)

	LOG	MATH	MAT	HOP	ELM	C-Index	HOT	SW	HOM
Stand. β -Gewicht	-0,238	-0,172	-0,078	0,188	-0,135	-0,129	-0,099	-0,069	-0,001
Konfidenzintervall (95 %)	-0,380 -0,089	-0,314 -0,029	-0,224 0,071	-0,069 0,450	-0,300 0,033	-0,239 -0,005	-0,311 0,129	-0,222 0,079	-0,148 0,156

Anmerkungen: Gemittelte β -Gewichte nach 1000 Bootstrapping-Durchläufen

Vordiplomnote mit der höheren Ausprägung dieser Skala. Entsprechendes zeigte sich auch schon bei der klassischen Kreuzvalidierung (siehe Tabelle 7.3).

Inkrementelle Validität über Abiturnote hinaus

Die bisherigen, hier vorgestellten Validierungsanalysen haben zur Prädiktion der Studiennoten lediglich das Self-Assessment verwendet. Im Folgenden wird darüber hinaus gehend geprüft, welche Verbesserung der Prädiktion über die Schulnoten hinaus durch das Self-Assessment erreicht werden kann. Die reine Abiturdurchschnittsnote erzielt in der reduzierten Stichprobe (wieder $n = 91$) – jeweils für sich betrachtet – eine Korrelation mit den Studiennoten als Kriterium von $r = 0,304$, die kognitiven Tests mit demselben Kriterium von $R = 0,445$ und die nicht-kognitiven Fragebogen von $R = 0,417$.

Betrachtet man alle Prädiktoren zusammen, wird eine multiple Korrelation von $R = 0,595$ erreicht (siehe Tabelle 7.5), wobei jeder der drei Prädiktoren eine signifikante Verbesserung der Varianzaufklärung zeigen. Bei der klassischen Kreuzvalidierung korrelieren die vorhergesagten Studiennoten mit den realen zu $R = 0,426$.

Tabelle 7.5
Vorhersage der mittleren Vordiplomnote im Hauptfach Informatik über das Abitur hinaus

Beschreibung	R	R^2	Korr. R^2	F	df	Signifikanz
Abiturdurchschnittsnote	0,304	0,092	0,082	9,066	1/89	0,003
+ kognitive Bestandteile	0,500	0,250	0,214	7,152	4/86	< 0,001
+ nicht-kognitive Bestandteile	0,595	0,354	0,273	4,387	10/80	< 0,001

Anmerkungen: $n = 91$; in der Spalte df werden zuerst die Freiheitsgrade für die erklärte, dann für die nicht erklärte Streuung aufgeführt.

Zur Absicherung wird auch hier ein Bootstrapping mit 1 000 Zufallsziehungen durchgeführt. Dabei zeigt sich, dass die zusätzliche Aufnahme der Abiturdurchschnittsnote zu den kognitiven und nicht-kognitiven Bestandteilen des Self-Assessments die mittlere multiple Korrelation von $R = 0,530$ um 10,0 % auf $R = 0,583$ erhöht. Das 95 %-Konfidenzintervall liegt zwischen $R = 0,487$ und $0,677$, das Bestimmtheitsmaß zwischen $0,237$ und $0,458$. Das Bootstrappingverfahren hilft auch, den besten unter den drei Prädiktoren (Abitur, kognitive bzw. nicht-kognitive Tests) herauszufinden. Dabei schwankt das Konfidenzintervall für die Korrelation der Abiturdurchschnittsnote mit den Studiennoten zwischen $r = 0,245$ und $0,516$ (korrigiertes Mittel $r = 0,388$). Betrachtet man hingegen die Einzelnoten Mathematik, Englisch, Physik und Informatik zusätzlich zur Gesamtnote und erreicht eine optimale Gewichtung mithilfe der multiplen Regression, so liegt das Konfidenzintervall zwischen $R = 0,320$ und $0,558$ (korrigiertes Mittel

$R = 0,444$). Somit kann in diesem Fall die Aufnahme der Einzelnoten statt der Durchschnittsnote die statistische Prädiktion verbessern. Die kognitiven Verfahren zeigen – wie bereits berichtet – eine multiple Korrelation zwischen $R = 0,352$ und $0,563$ (im Mittel $R = 0,464$), die nicht-kognitiven zwischen $R = 0,260$ und $0,477$ (im Mittel $R = 0,368$). Die kognitive Testbatterie kann entsprechend dieser Validierung mehr Varianz im Kriterium aufklären als die Abiturdurchschnittsnote.

Receiver Operating Characteristic

Die Basis der folgenden Analysen ist wiederum die Signalentdeckungstheorie. Die eigentlich aus der Elektrotechnik stammende Theorie wird in zahlreichen Studien auch in der Medizin verwandt, um die Sensitivität beziehungsweise Spezifität eines diagnostischen Tests zu eruieren. Aber auch für die Prognose des Studienerfolgs lassen sich insbesondere *ROC-Kurven* (*Receiver Operating Characteristic*) verwenden, um bei einem binären Kriterium (hier der Studienerfolg) die Validität verschiedener Prädiktoren anschaulich darzustellen (siehe Abschnitt 6.3.4).

Das in diesem Kontext zu entdeckende „Signal“ ist der Erfolg im Studium. Da in dieser Studie aus Gründen des Datenschutzes keine Studienabbrecherdaten vorliegen, wird ein Mediansplit der Gruppe hinsichtlich der Vordiplomnote vorgenommen: Personen, die eine mittlere Vordiplomnote von 4,0 oder schwächer aufzeigen, werden zu den bisher nicht erfolgreichen Studenten gezählt. Als Prädiktoren, die das entsprechende Signal entdecken sollen, werden sowohl die gesamte Testbatterie als auch die Abiturdurchschnittsnote verwandt. Die allgemeine Arbeitshypothese (H_1) geht davon aus, dass eine Person geeignet ist. Die Nullhypothese (H_0) nimmt hingegen an, dass eine Person nicht geeignet ist.

In der Abbildung 7.4 lassen sich die Prognose- beziehungsweise Entdeckungsfähigkeiten der Abiturdurchschnittsnote, der Testbatterie und der Testbatterie mit gemeinsamer Berücksichtigung der Abiturnote vergleichen.

Bei den letzten beiden Prädiktoren wurden die Gewichte aus den bereits berichteten multiplen Regressionen verwandt. Zur Analyse wurde die gesamte Stichprobe $n = 182$ der Studienanfänger der Diplom-Informatik herangezogen. So zeigt sich weiter, dass die Testbatterie im direkten Vergleich zur Abiturdurchschnittsnote der wichtigere Prädiktor für das binäre Kriterium des Studienerfolgs ist. Die Fläche unter der ROC-Kurve ist entsprechend bei der alleinigen Verwendung der Testbatterie ($AUC = 0,759$) größer als bei der Verwendung der Abiturnote ($AUC = 0,705$). Dieser Effekt gilt für nahezu alle Bereiche der Skala. Eine weitere Verbesserung der Prädiktion und somit auch eine Vergrößerung der Fläche zeigt sich bei der gemeinsamen Verwendung von Abitur und Testbatterie ($AUC = 0,795$). Diese Aussage gilt für die gesamte Skala. Das entsprechende 95 %-Konfidenzintervall liegt zwischen $AUC = 0,730$ und $0,860$.

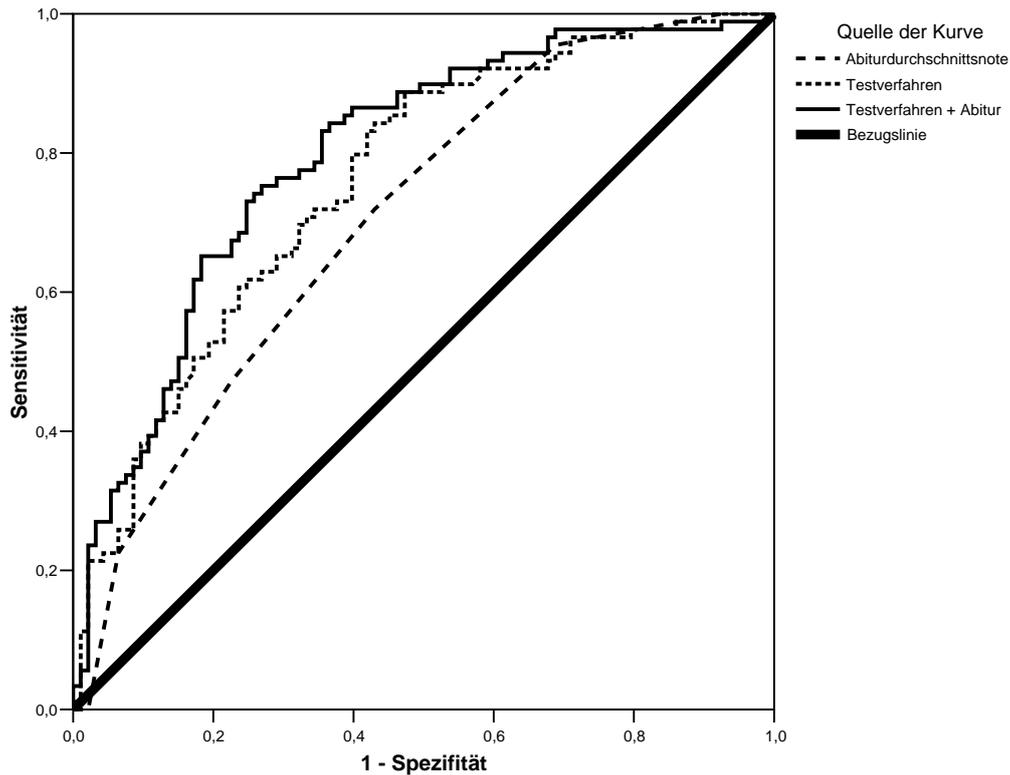


Abbildung 7.4: Sensitivität und Komplement der Spezifität der Prädiktoren für die Vorhersage der späteren Studiennoten

7.4.3 Vorhersage der Anzahl bestandener Prüfungen im Hauptfach Informatik

Die bereits berichteten Regressionsanalysen dienen der Vorhersage der mittleren Vordiplomnote. Ein wesentliches Erfolgskriterium ist neben den Studiennoten auch die Dauer des Studiums, die eine enge Verbindung zum Konstrukt der Procrastination aufweist (Helmke & Schrader, 2000). Aufgrund der Bedeutung eines zügigen Studiums für den Studenten selber, aber eben auch für die Institution und für die Volkswirtschaft wird der Studienfortschritt hier ebenfalls erfasst, und zwar operationalisiert über die Anzahl der bestandenen Vordiplomprüfungen. Als *nicht bestanden* werden Prüfungen klassifiziert, die mit *mangelhaft* beurteilt wurden. Lediglich Vordiplomprüfungen wurden berücksichtigt, Semestralprüfungen hingegen nicht. Alle hier einbezogenen Personen gehören zur selben Kohorte Studierender, die im Wintersemester 2004/2005 mit der Programmierungsvorlesung im ersten Fachsemester ihr Studium der Informatik im Hauptfach aufgenommen haben.

Von der betrachteten Gesamtstichprobe von $n = 182$ haben sich 19 Personen in dem Betrachtungszeitraum für eine, 28 zwei, 39 drei, 50 vier, 45 fünf und eine Person insgesamt für sechs Prüfungen angemeldet. Entsprechend liegt der Mittelwert der geplanten Prüfungen in der Stichprobe bei 3,423. Davon haben 43 Personen in dem Betrachtungszeitraum keine Prüfung

bestanden, 53 eine, 28 zwei, weitere 28 drei, wiederum 28 vier und lediglich zwei Personen insgesamt fünf Prüfungen. Entsprechend liegt der Mittelwert bestandener Prüfungen in der Stichprobe bei 1,731.

Kriteriumsvalidierung mit Kreuzvalidierung

In der ersten Hälfte der Stichprobe zeigt sich eine Varianzaufklärung der kognitiven Tests von 12,8 % ($R = 0,358$) (siehe Tabelle 7.6). Die Hinzunahme der nicht-kognitiven Fragebogen erhöht die Aufklärung gemeinsamer Varianz auf 24,8 % ($R = 0,498$). Diese Steigerung ist bei einer Irrtumswahrscheinlichkeit von $\alpha = 0,05$ eine signifikante Verbesserung der Prognosefähigkeit. Im direkten Vergleich der kognitiven Tests und der nicht-kognitiven Fragebogen als alleiniger Prädiktor fällt auf, dass nicht-kognitiven in der hier betrachteten Stichprobe tendenziell mehr als die kognitiven Tests aufklären, nämlich 14,4 % ($R = 0,379$) statt 12,8 % ($R = 0,358$). Dieser Unterschied ist aber nicht bedeutsam.

Tabelle 7.6

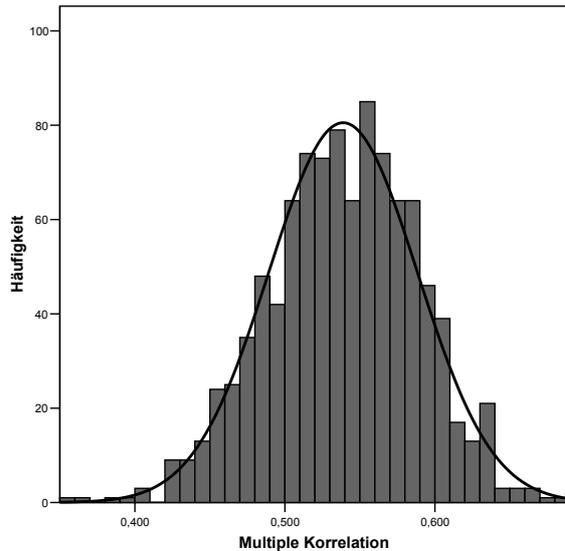
Multiple Regression zur Vorhersage der Anzahl bestandener Vordiplomprüfungen im Hauptfach Informatik

Beschreibung	R	R^2	Korr. R^2	F	df	Signifikanz
Kognitive Bestandteile	0,358	0,128	0,098	4,259	3/87	0,007
+ nicht-kognitive Bestandteile	0,498	0,248	0,165	2,971	9/81	0,004

Anmerkungen: $n = 91$; in der Spalte df werden zuerst die Freiheitsgrade für die erklärte, dann für die nicht erklärte Streuung aufgeführt.

Zuerst wird wieder eine klassische Kreuzvalidierung durchgeführt: Während sich in der ersten Stichprobe ein Korrelation der vorhergesagten und der realen Prüfungsanzahl von $R = 0,498$ gezeigt hat, liegt diese in der zweiten zur Kreuzvalidierung verwandten Stichprobe niedriger ($R = 0,442$).

Vergleichbar mit dem oben beschriebenen Vorgehen wird auch für das Kriterium der Prüfungsanzahl ein Bootstrapping mit 1000 Datensätzen durchgeführt: Das 95 %-Konfidenzintervall für die multiple Korrelation liegt dabei zwischen $R = 0,439$ und $0,636$ (im Mittel $R = 0,541$), für das Bestimmtheitsmaß zwischen $0,193$ und $0,404$ (siehe Abbildung 7.5). Die Varianzaufklärung der kognitiven Tests als auch der nicht-kognitiven Fragebogen liegt im Mittel bei $R^2 = 0,177$. Tabelle 7.7 lässt die Vorhersagekraft der einzelnen Subtests und Fragebogen genauer analysieren. Die beste Varianzaufklärung erreichen diesmal die Handlungsorientierung bei Tätigkeitsinitiierung (*HOT*), der Logiktest (*LOG*) und das Interesse (*INT*). Bei diesen drei schließt das Konfidenzintervall für die β -Gewichte den Null-Wert nicht mit ein. Auffällig ist zudem, dass sich wieder ein negatives β -Gewicht bei der *HOP*-Skala (siehe 5.3.2) zeigt. Bei allen anderen Skalen sind die Gewichte positiv, so dass dort die linearen Regressionsbeiträge stetig steigen.



Beschreibung	R
Mittelwert (korrigiert)	0,541
Konfidenzintervall (95-%)	0,439 0,636

Abbildung 7.5: Verteilung der multiplen Korrelation zwischen den Prädiktoren und der Anzahl der bestanden Prüfungen. Ermittelt mithilfe von 1 000 Bootstrapping-Datensätzen. Die Korrelationen sind vor der Mittelung Fisher-z-transformiert worden.

Tabelle 7.7
Standardisierte β -Gewichte zur Vorhersage der Prüfungsanzahl

	<i>LOG</i>	<i>MATH</i>	<i>MAT</i>	<i>HOT</i>	<i>HOP</i>	<i>ELM</i>	<i>C-Index</i>	<i>HOM</i>	<i>SW</i>
Stand. β -Gewicht	0,264	0,116	0,027	0,287	-0,278	0,160	0,154	0,069	0,035
Konfidenzintervall	0,089	-0,033	-0,112	0,093	-0,501	-0,005	0,035	-0,073	-0,130
	0,433	0,263	0,181	0,484	-0,028	0,308	0,273	0,213	0,214

Anmerkungen: Gemittelte β -Gewichte nach 1 000 Bootstrapping-Durchläufen

Inkrementelle Validität über Abiturnote hinaus

Tabelle 7.8 zeigt die Änderung der Varianzaufklärung, wenn bei der Regressionsanalyse neben den Beratungstests auch die Abiturnote berücksichtigt wird. Somit kann überprüft werden, ob das Self-Assessment die Vorhersage noch erhöhen würde, wenn das Abitur bereits zur Prädiktion verwendet wird. Ein wesentliches Ergebnis ist, dass die Abiturnote allein betrachtet die geringste Korrelation mit dem Kriterium erzielt ($r = 0,296$). Die nicht-kognitiven Bestandteile erhöhen den statistischen Zusammenhang um 54,7 % ($R = 0,458$). Auch über die beiden Prädiktoren hinaus zeigt sich eine signifikante Verbesserung der Vorhersage durch die kognitiven Scores ($R = 0,538$). Bei einer Kreuzvalidierung zeigt sich statt der berichteten multiplen Korrelation von $R = 0,538$ eine Korrelation von $R = 0,522$.

Beim Bootstrapping lässt sich eine Korrelation zwischen allen Prädiktoren und dem Studienerfolg zwischen $R = 0,498$ und $0,678$ (Konfidenzintervall bei einer Irrtumswahrscheinlichkeit von 5 %, im Mittel $R = 0,590$) berechnen. Die Abiturgesamtnote allein korreliert im korrigierten Mittel mit dem Kriterium zu $R = 0,385$ (Konfidenzintervall: $0,248 \leq R \leq 0,507$), die kognitiven

Tabelle 7.8

Multiple Regression zur Vorhersage der Anzahl bestandener Vordiplomprüfungen im Hauptfach Informatik

Beschreibung	R	R^2	Korr. R^2	F	df	Signifikanz
Abiturnote	0,296	0,088	0,077	8,535	1/89	0,004
+ nicht-kognitive Bestandteile	0,458	0,210	0,143	3,152	7/83	0,005
+ kognitive Bestandteile	0,538	0,290	0,201	3,265	10/80	0,001

Anmerkungen: $n = 91$; in der Spalte df werden zuerst die Freiheitsgrade für die erklärte, dann für die nicht erklärte Streuung aufgeführt.

Tests zu $R = 0,418$ (Konfidenzintervall: $0,295 \leq R \leq 0,532$), ebenso wie die nicht-kognitiven Tests ($R = 0,418$; Konfidenzintervall: $0,290 \leq R \leq 0,522$). Somit zeigen sowohl die kognitiven als auch die nicht-kognitiven Skalen eine bessere Varianzaufklärung im Kriterium *Prüfungsanzahl* als die Abiturdurchschnittsnote.

7.4.4 Vorhersage der Studiennoten im Nebenfach Informatik

Die hier betrachtete Stichprobe besteht aus 412 Studenten im Nebenfach Informatik. Weitere Informationen zu den Teilnehmern an dieser Studie können dem Abschnitt 7.3 entnommen werden. Für die Vorhersage der Studiennoten im Nebenfach Informatik werden die β -Gewichte und Konstanten der Regressionsanalyse für Hauptfachstudenten herangezogen. Diese Kreuzvalidierung gelingt in einem ähnlichen Maße wie das oben beschriebene Vorgehen: Die Ergebnisse der Testbatterie korrelieren allein mit $r = 0,347$ mit den Studiennoten (statt $R = 0,562$ bei der ersten Stichprobenhälfte), mit einer Kombination von Test und Abitur mit $r = 0,347$ (statt $R = 0,595$).

Das Konfidenzintervall für die multiple Korrelation – berechnet im Bootstrapping-Verfahren – liegt bei einer Irrtumswahrscheinlichkeit von 5 % zwischen 0,422 und 0,568, im korrigierten Mittel bei $R = 0,499$ (siehe Tabelle 7.9). Bei einer Berücksichtigung der Abiturnote verschiebt sich das Konfidenzintervall auf $R = 0,469$ und 0,614 (korrigiertes Mittel 0,543). Somit verbessert die Aufnahme der Abiturnote die Prognosefähigkeit über die Testbatterie hinaus.

7.4.5 Vorhersage der Anzahl bestandener Prüfungen im Nebenfach Informatik

Im Mittel haben die Nebenfachstudenten 0,888 Prüfungen bestanden beziehungsweise im Median 1,000. 123 Studenten haben keine, 245 eine Prüfung und 44 mehr als eine Prüfung bestanden. In der Ausgangsstichprobe der Kreuzvalidierung korrelieren die Ergebnisse der Testbatterie zu $R = 0,498$ mit der Anzahl bestandener Prüfungen. In der klassischen Kreuzvalidierung zeigt

sich eine Korrelation zwischen der durch den Test prognostizierten Anzahl der Prüfungen und der realen Anzahl von $r = 0,198$. Im Fall der Berücksichtigung des Abiturs verbessert sich die Vorhersage auf $r = 0,232$ (statt $R = 0,538$ in der Ausgangsstichprobe der Kreuzvalidierung).

Tabelle 7.9

Mittlere multiple Korrelation zur Vorhersage der Vordiplomnoten und Anzahl bestandener Prüfungen im Hauptfach und Nebenfach Informatik

Kriterium	Prädiktor	Hauptfach			Nebenfach		
		R	95-%-KI		R	95-%-KI	
Studiennote	Test	0,530	0,434	– 0,622	0,499	0,422	– 0,569
	Abitur + Test	0,583	0,487	– 0,677	0,543	0,469	– 0,614
Prüfungsanzahl	Test	0,541	0,439	– 0,636	0,303	0,227	– 0,383
	Abitur + Test	0,590	0,498	– 0,678	0,324	0,243	– 0,402

Anmerkungen: Jeweilige Ergebnisse der Bootstrapping-Prozeduren mit 1000 Stichprobenziehungen. Die Korrelationen sind Fisher-z-transformiert gemittelt worden.

Im Rahmen des Bootstrappings zeigen sich folgende Werte: Das Konfidenzintervall der Korrelation zwischen der Testbatterie und des Kriteriums liegt bei $0,227 \leq R \leq 0,383$ (korrigiertes Mittel 0,303). Wird das Abitur noch zusätzlich als Prädiktor aufgenommen, so liegt das Intervall bei $0,243 \leq R \leq 0,402$ (korrigiertes Mittel 0,324). Das Abitur kann also auch hier die Vorhersage bedeutsam verbessern.

Die Tabelle 7.9 zeigt die wesentlichen Ergebnisse des Bootstrappings in der Übersicht.

7.5 Diskussion

In Studie 3 wurde die Kriteriumsvalidität des Self-Assessments für den Studiengang Diplom-Informatik und für das Nebenfach Informatik überprüft. In den Analysen wurden unterschiedliche Definitionen von Studienerfolg verwandt, um zwei (teilweise gegenläufige) Aspekte des Lern- und Prüfungsverhaltens erfassen zu können: So wurden sowohl die mittlere Vordiplomnote als auch die Anzahl bereits bestandener Prüfungen berücksichtigt.

Es konnte in dieser Studie aufgezeigt werden, dass auch durch ein freiwilliges und webbasiertes Beratungssystem mit kurzer Testdauer eine gute Prädiktion des Studienerfolgs in der ersten Studienhälfte erreicht werden kann – und zwar bereits vor dem Abitur: Die kognitiven Tests im Self-Assessment korrelieren im Vergleich aller Prädiktoren am höchsten mit den Vordiplomnoten im Hauptfach (mittleres $R = 0,464$, alle berichteten Werte sind unkorrigiert). Die Abiturdurchschnittsnote klärt etwas weniger Varianz auf ($r = 0,388$). Die nicht-kognitiven Fragebogen korrelieren zwar geringer als diese beiden Prädiktoren mit dem Kriterium, aber auch

noch in einer bedeutsamen Größenordnung ($R = 0,368$). Verwendet man das Self-Assessment komplett, so zeigt sich eine multiple Korrelation von $R = 0,530$ (das 95 %-Konfidenzintervall liegt zwischen $R = 0,434$ und $0,622$). Alle drei Prädiktorgruppen, also Tests, Fragebogen und Abitur, verbessern inkrementell die Varianzaufklärung. Zusammen korrelieren die Prädiktoren mit $R = 0,583$ mit der mittleren Vordiplomnote. Im Rahmen der Studie wurden die Prädiktoren über eine multiple Regression möglichst gut linear kombiniert. Die Wahrscheinlichkeit, dass zufällige Zusammenhänge zwischen Kriterium und Prädiktor die Ergebnisse beeinträchtigen, wurde hingegen durch die Bootstrapping-Prozedur – also durch die Auswahl und Gewichtung einzelner Teilnehmer-Daten – minimiert.

Die Publikation von Meyer et al. (2001) hilft dabei, diese hier berichteten Korrelationskoeffizienten besser einschätzen zu können: Die Autoren haben statistische Beziehungen aus vielen Bereichen, u.a. auch aus der Medizin und Psychologie, gesammelt und erlauben damit einen anschaulichen Vergleich zwischen den eigenen und in den anderen Studien berichteten Größen: Der Zusammenhang zwischen Gewicht und Körpergröße wird dort mit $r = 0,400$ angegeben, die Korrelation zwischen Geschlecht und Armkraft mit $r = 0,550$, die zwischen Rauchen und Lungenkrebs mit $r = 0,08$ und die zwischen der Gabe von Nikotin-Pflastern und Raucherentwöhnung mit $r = 0,18$. Anhand dieser Werte zeigt sich anschaulich, wie vergleichsweise valide die Testbatterie den Studienerfolg in den ersten Semestern vorhersagen kann (Korrelation von $r = 0,530$ mit der Studiennote bzw. $r = 0,541$ mit der Prüfungsanzahl) – trotz der doch kurzen, webbasierten Testung. Allerdings muss dabei wieder – wie bereits im vorigen Absatz erwähnt – auf die Besonderheiten der multiplen Regression eingegangen werden, die die Gewichtung der unterschiedlichen Skalen optimiert vornimmt, ohne unbedingt inhaltlich nachvollziehbar zu sein. Daher kann die Tabelle 7.1 hier zur Einschätzung weiterhelfen, indem sie Einzelkorrelationen aufführt: Dort sind unkorrigierte Korrelationskoeffizienten berichtet worden, die schon alleine, zum Beispiel im Fall des Logiktests ($r = 0,404$), eine hervorragende Validität aufzeigen.

Für eine Prognose der Anzahl bestandener Prüfungen im Vordiplom zeigen sich Werte in einer vergleichbaren Größenordnung: Die Abiturgesamtnote alleine korreliert im korrigierten Mittel mit dem Kriterium zu $r = 0,385$, sowohl die kognitiven Tests als auch die nicht-kognitiven Fragebogen zu $R = 0,418$. Somit klären hier beide Prädiktoren des Self-Assessments für sich mehr Varianz auf als die Abiturdurchschnittsnote. Zudem zeigen sowohl die kognitiven Tests als auch die nicht-kognitiven Fragebogen eine gleich hohe statistische Beziehung zum Kriterium. Die Bedeutung von Motivation und Volition für die Anzahl bestandener Prüfungen fällt höher aus als bei der reinen Betrachtung der Studiennote (siehe oben). Gleichzeitig sinkt die Varianzaufklärung durch die kognitiven Tests leicht im Vergleich zur Prädiktion der Studiennote. Alle Prädiktoren korrelieren gemeinsam mit $R = 0,590$ mit der Prüfungsanzahl und damit im nahezu identischen Umfang wie mit der Vordiplomnote.

In dieser Kriteriumsvalidierung konnte gezeigt werden, dass die Testbatterie in der hier beschriebenen Stichprobe ein besserer Prädiktor für die Vordiplomnote und auch für die Zahl bestandener Prüfungen ist als die Abiturdurchschnittsnote.

Auf Ebene der einzelnen Tests beziehungsweise Fragebogen verglichen zeigt der Logiktest die höchste Validität bei dem Kriterium der *Studiennoten*. Das entsprechende 95 %-Konfidenzintervall des β -Gewichts umschließt (genauso wie übrigens auch das von *MATH* und *C-Index*) den Null-Wert nicht. Die größte Varianzaufklärung unter den nicht-kognitiven Bestandteilen des Self-Assessments erreichen der Fragebogen zur extrinsischen Lernmotivation *ELM*, der Fragebogen zur Handlungsorientierung *HOP* und der Interessenfragebogen *INT* (hier operationalisiert als *C-Index*).

Die Skala *HOP* (Handlungsorientierung bei der Handlungsplanung) zeigt in der Studie nach dem Logiktest *LOG* das zweitgrößte, allerdings positive β -Gewicht. Dies deutet darauf hin, dass regressionsanalytisch bei einem höheren *HOP*-Wert vor dem Studium eher schwächere Studienleistungen wahrscheinlich sind. Die anderen beiden Skalen der Handlungsorientierung zeigen, genauso wie alle anderen Skalen auch, negative β -Gewichte. Aufgrund der gegenläufigen Skalierung von Studiennoten und Testergebnis ist dieses Ergebnis erwartungskonform. Aus welchem Grund die Skala *HOP* regressionsanalytisch bei beiden Studienerfolgskriterien positiv gewichtet wurde, muss weiter untersucht werden. Dieses Ergebnis lässt sich scheinbar nicht mit der Theorie nach Kuhl (1987) in Einklang bringen. Mehr Forschung mit einer zweiten Stichprobe und eventuell nach einer erneuten Revision der Aufgaben scheint nötig.

Für die Vorhersage der Anzahl bestandener Prüfungen sind die nicht-kognitiven Bestandteile des Self-Assessments tendenziell wichtiger als für die Prädiktion der Studiennoten: So zeigt die Skala Handlungsorientierung bei Tätigkeitsinitiierung *HOT* die höchsten β -Gewichte aller Test- und Fragebogenskalen, direkt gefolgt von der Handlungsorientierung bei Planung und dem Logiktest. Je nach gewähltem Kriterium für den Studienerfolg verschiebt sich also die Prädiktionskraft der einzelnen Self-Assessment-Bestandteile. So ist im Fall der Prüfungsanzahl die Varianzaufklärung zwischen den Tests und den Fragebogen ähnlich, nicht aber beim Kriterium der Studiennoten. Gerade die hohe Bedeutung der Skala Handlungsorientierung nach Tätigkeitsinitiierung deutet darauf hin, dass die Anzahl bestandener Prüfungen nicht nur durch Fähigkeitsaspekte (wie zum Beispiel dem logischen Schlussfolgern) vorhersagbar ist. Aspekte, die getätigte Planung (*HOT* auch in Wirklichkeit umzusetzen, stellen dieser Studie zufolge ein wichtiges Element des Studienerfolgs dar.

Bis hierhin wurden nur die Ergebnisse der Hauptfach-Stichprobe näher betrachtet. Die Nebenfach-Stichprobe ist im direkten Vergleich zu dieser heterogener: Die Studenten studieren im Hauptfach komplett unterschiedliche Fächer, haben daher eventuell auch etwas andere Eignungs- und Neigungsprofile. Da die einzelnen Prüfungsordnungen der Fächer zudem für

das Nebenfach Informatik im Detail unterschiedliche Prüfungen vorsehen, sind auch die einzelnen mittleren Studienfachnoten nur eingeschränkt miteinander zu vergleichen. Sinnvoll wäre es daher, bei einer größeren Stichprobengröße die einzelnen Gruppen von Nebenfach-Studenten getrennt zu betrachten. Dennoch zeigt Tabelle 7.9, dass eine Prognose bei der Prüfungsnoten mit $R = 0,499$ ähnlich gut funktioniert wie bei der Hauptfach-Stichprobe ($R = 0,530$). Die statistische Beziehung zum Kriterium *Prüfungsanzahl* ist aber mit $R = 0,303$ vergleichsweise gering. Scheinbar zeigt sich dort der Einfluss der unterschiedlichen Prüfungsordnungen der Fächer stärker als bei der mittleren Vordiplomnote.

Alle hier genannten Korrelationskoeffizienten sind unkorrigiert. Da davon auszugehen ist, dass das Kriterium (hier zum Beispiel die Studiennoten) bereits unreliabel ist, wird alleine dadurch die statische Beziehung zwischen dem Test und dem Kriterium gemindert, ohne dass dies vom Test verursacht wird. An dieser Stelle wird aus diversen Gründen davon abgesehen, diesen Einfluss zu korrigieren: Erstens erscheint es schwierig, die Reliabilität der Studiennoten abzuschätzen, um überhaupt eine Korrektur durchführen zu können. Empirische Studien zur Reliabilität von Noten gibt es in Deutschland lediglich aus dem schulischen, nicht aus dem universitären Bereich. Für den US-amerikanischen Bereich ließe sich auf Kuncel, Hezlett und Ones (2001) zurückgreifen, die von internen Konsistenzen der College-Noten zwischen $r_{tt} = 0,80$ und $r_{tt} = 0,83$ berichten. Ob in Deutschland Studiennoten ähnlich reliabel sind, bleibt hingegen unklar, da die Hochschulsysteme und die Systeme der Notenvergaben zu unterschiedlich sind. Zudem ist eine Korrektur des Kriteriums, also hier der Studiennoten, praxisfern, da in der Realität eben diese Korrektur nicht greift und die Varianzaufklärung der Tests auf jeden Fall gemindert wird, auch wenn die Tests selbst nicht Verursacher sind, sondern die Kriterien. Eine weitere, häufig durchgeführte Korrektur beschäftigt sich mit der eingeschränkten Varianz durch Auswahlprozesse. Falls hier Daten von einem Auswahlverfahren für Studieninteressierte analysiert werden würden und mit dem Verfahren bereits die Hochschulzulassung geregelt wäre, so wären die Varianz der später von den Studenten erreichten Studiennoten eingeschränkt und die Kriteriumsvalidität dadurch gemindert. So ist anzunehmen, dass Personen, die aufgrund des Tests abgelehnt wurden, auch schwächere Studienleistungen erbracht hätten, hätte man sie denn zugelassen. Die Kriteriumsvalidierung hier wurde allerdings so durchgeführt, dass keine Varianzeinschränkung in der Studiennote vorliegt: Erstens habe alle Personen am Self-Assessment teilgenommen. Zudem wurde keine Personen von einem Studium ausgeschlossen, und von allen Personen wurden die mittleren Studiennoten ausgewertet. Eine Korrektur der Varianzeinschränkung ist hier also nicht nötig.

Im Folgenden sollen mögliche Einschränkungen dieser Studie näher beleuchtet werden. Zudem sollen Lösungsmöglichkeiten angeführt werden, die in vergleichbaren Studien berücksichtigt werden können:

Fraglich ist, welche Relevanz die hier berichteten Prädiktor-Kriterium-Beziehungen über das beschriebene Self-Assessment hinaus haben. Auch wenn sich in dieser Studie zeigen konnte, dass nicht nur die Studierfähigkeit, sondern eben auch die Motivation und Volition die Vorhersage bedeutsam unterstützten, ist diese Information nicht unmittelbar für alle Hochschulen, Fächer und Projekte übertragbar. Es bleibt ungeprüft, ob die hier erzielten Ergebnisse (zum Beispiel die hohen Prognosefähigkeiten der nicht-kognitiven Fragebogen) eng mit den Besonderheiten des Studiums der Informatik an der RWTH Aachen verbunden sind. So wäre es denkbar, dass die Skalen der Motivation und Volition in bereits auf Bachelor umgestellten Studiengängen oder an Fachhochschulen statt Technischen Hochschulen weniger relevant sind als hier beschrieben. Es wäre daher durchaus sinnvoll, die Prognosefähigkeiten der hier vorgestellten Self-Assessment-Batterie auch in anderen vergleichbaren Studienfächern an der RWTH Aachen beziehungsweise an anderen Hochschulen zu überprüfen. Somit wäre nachweisbar, ob die Ergebnisse (zum Beispiel die bessere Prognosefähigkeit des Self-Assessments im Vergleich mit der Abiturdurchschnittsnote) über mehrere Stichproben unterschiedlicher Fächer hinweg übertragbar sind.

Noch allgemeiner kann hier ebenfalls angeregt werden, über die Relevanz hinsichtlich der Studierendenauswahl nachzudenken. Belegt ist, dass Personen auf Items der Persönlichkeit und des Interesses im Rahmen von Orientierungsverfahren ehrlicher antworten, als sie dies bei Auswahlprozessen von Hochschulen tun würden. Das bereits erwähnte *Cheating* und die *soziale Erwünschtheit* (siehe Abschnitte 2.5 und 5.3.1) führt eventuell dazu, dass zwar eine hohe Validität auch der nicht-kognitiven Faktoren resultiert, diese Ergebnisse aber nicht für die Studierendenauswahl generalisierbar sind. Da derzeit keine objektiven Persönlichkeitstests mit nachweislich hohen Kriteriumsvaliditäten für entsprechende Anforderungsbereiche existieren, bleibt es ein Vorteil der vorgestellten Self-Assessments, dass sie auch die Motivations- und Volitionsbereiche prüfen. Umso wichtiger scheint es, Self-Assessments und Verfahren der Hochschulauswahl nicht als Gegenspieler, sondern als gegenseitige Förderer zu sehen.

Aufgrund des Datenschutzes war es in dieser Studie nicht möglich, Daten der Studiennoten und des Self-Assessments mit anderen Daten des Studierendensekretariats zusammenzuführen. Auch war es nicht möglich zu identifizieren, welche Personen zwar am Self-Assessment im Rahmen des Informatikstudiums teilgenommen, später aber keine Anmeldung zu einer Prüfung eingereicht haben. Diese Informationen wären hingegen wichtig, wenn man das Studienerfolgskriterium *Studienabbruch* untersuchen möchte. Demgemäß mussten die vorliegenden Analysen auf Studiennoten und die Anzahl bestandener Prüfungen beschränkt bleiben. Die Prozedur der Anonymisierung sollte im nächsten Ausbauschritt entsprechend erweitert werden, so dass ebenfalls Exmatrikulationsinformationen aus dem Studierendensekretariat ausgewertet werden können. Zusätzlich sollte dabei der Grund für einen Studienabbruch angegeben werden, damit Studienortwechsler nicht als Studienabbrecher geführt werden.

Anders als die Prüfungsnoten an der Hochschule wurden die Abiturdurchschnittsnoten in einer Selbstauskunft erfragt. Aus diesem Grund muss hier diskutiert werden, ob die Validität der Schulnoten geringer ausgefallen ist, weil die Noten nicht überprüft wurden (zum Beispiel über einen Datenabgleich mit dem Studierendensekretariat) oder weil diese keine bessere Vorhersagekraft – zumindest für diese Stichprobe hier – besitzen. Bei US-amerikanischen Studien haben Kuncel, Crede und Thomas (2005) für selbst berichtete Noten im Vergleich zu direkt von der Schule erhobenen eine geringere Konstrukt- und Kriteriumsvalidität nachgewiesen, gerade für Personen mit schwächeren Schulnoten. Dickhäuser und Plenter (2005) haben im Rahmen der TIMS-Studie in Deutschland eine ähnliche Untersuchung durchgeführt. Es werden bei beiden Studien hohe Korrelationen zwischen den selbst und fremd berichteten Noten genannt, die selbst berichteten Noten fielen im Mittel allerdings etwas besser aus. Die Korrelationen sowohl der selbst berichteten wie der vom Lehrer angegebenen Note mit der über die TIMSS-Aufgaben erfassten Leistung unterscheiden sich statistisch nicht voneinander. Zudem fällt diese Verzerrung in der deutschen Studie unabhängig von Geschlecht, mathematischem Fähigkeitsselfkonzept und mathematischer Leistung aus. Für die Behauptung unterschiedlicher Kriteriumsvalidität selbst berichteter und fremd berichteter Noten ergibt sich damit – zumindest für den deutschen Bereich – keine Unterstützung. Daher können aus der Forschung keine Anhaltspunkte dafür gefunden werden, dass die Ergebnisse nur aufgrund einer verfälschten Notenwiedergabe entstanden sind.

Zusätzliche Forschung sollte betrieben werden, um das Kriterium Studiennoten noch umfassender analysieren zu können. In der Studie 3 wurden die bereits erreichten gemittelten Studiennoten als Kriterium verwandt. Sinnvoll wäre es, die abschließende Vordiplomnote und die Diplomnote auszuwerten und als weitere Kriterien zu verwenden. Die bisherige Literatur deutet aber darauf hin, dass Vordiplom- und Diplomnoten hoch korrelieren (Gold & Souvignier, 1997).

Gerade für Daten, die im Rahmen von Internettestungen entstanden sind, ist die Datenqualität und das Vorgehen beim Data-Screening zu diskutieren: Im Rahmen der oben genannten Studien wurden Datensätze – wie bereits berichtet – konservativ anhand der Antwortzeiten eliminiert. Bei kognitiven Leistungstests hat sich dieses Vorgehen bewährt. Die im Pretest gefundenen niedrigsten Lesezeiten wurden als Mindestbearbeitungszeit angenommen (siehe Abschnitt 5.3.1). Für nicht-kognitive Fragebogen erscheint dies nur wenig sinnvoll, besteht doch bei diesen zwischen der Lese- und der gesamten Bearbeitungszeit samt der Entscheidung für eine Antwort nur eine geringe Zeitdifferenz. Hier müssen zusätzliche Verfahren gesucht werden, die das Identifizieren nicht ernsthafter Antworten verbessern. Denkbar wäre es, für die Fragebogen eine Patternanalyse (Zimmerhofer & Hornke, 2005a) durchzuführen, die gleichförmige Antworten (zum Beispiel stets erste Antwortkategorie) auffindet. Das Identifizieren zufälliger Antwortmuster ist mit einer Konsistenzanalyse (Johnson, 2005) zu leisten.

Kapitel 8

Studie 4: Akzeptanz

8.1 Einleitung

Als erfolgreich kann ein Beratungssystem für Studieninteressierte selbst bei nachgewiesener Kriteriumsvalidität (siehe Abschnitt 7) nur dann gelten, wenn es von der Zielgruppe auch intensiv genutzt wird. Sollten nur wenige, zudem vielleicht eher besser informierte Personen das System zur Beratung nutzen, so würde das Ziel einer verbesserten Information, einer daraus resultierenden höheren Studienzufriedenheit und geringeren Studienabbruchquoten breiter Studieninteressenten-Gruppen verfehlt werden. Ein freiwilliges Beratungssystem, wie hier das Self-Assessment, lebt von der „Mund-zu-Mund-Propaganda“, also von Weiterempfehlungen an Freunde und Bekannte. Der Mehrwert einer Teilnahme muss entsprechend einsichtig und auch eingängig vermittelbar sein. Nur somit ist – auch ohne hohe Marketingausgaben – ein hoher Bekanntheitsgrad des Beratungssystems in der Zielgruppe und entsprechend die intendierte Wirkung zu erreichen.

Webbasierte Beratungssysteme – wie das hier beschriebene Self-Assessment – stellen zudem häufig den ersten Kontakt der Hochschulen zu Studieninteressierten her. Personen werden auf eine Hochschule aufmerksam und lernen indirekt (primär wollten sie ja mehr über eigene Stärken und Schwächen erfahren) die Vorteile des Studienganges beziehungsweise des Fachbereichs und der Hochschule kennen. Somit stellen Self-Assessments auch eine effiziente, da vergleichsweise kostengünstige Möglichkeit dar, einen Imagewandel der Hochschule zu forcieren.

Eine hohe Teilnehmerzahl zu erreichen, stellt allerdings lediglich einen Teilerfolg dar. Im Mittelpunkt der Betrachtung muss ferner stehen, ob Studieninteressierte die Rückmeldung bei der individuellen Studienfachwahl auch berücksichtigen und nicht nur ungelesen oder zumindest unberücksichtigt vergessen. Betrachtet wird also, ob die postulierte Steuerfunktion des Beratungssystems (siehe Abschnitt 2.5) auch wirklich in der Praxis existiert, also geeignete Personen nach einer Beratung das bereits vorher avisierte Studienfach wählen, weniger geeignete Personen sich hingegen für ein anderes Fach entscheiden. Wichtig dafür ist, dass die Teilnehmer das

System und die Rückmeldung akzeptieren (Kersting, 1998) und ihr Vertrauen schenken: Eine hohe Realitätsnähe, ein hoher wahrgenommener Anforderungsbezug und eine entsprechend wahrgenommene Augenscheinvalidität des Verfahrens sind wichtige Voraussetzungen für einen Erfolg des Systems.

Um die Beliebtheit des Systems zu eruieren, werden in dieser vierten Teilstudie daher sowohl die Anzahl der Besucher und Teilnehmer des Self-Assessments als auch die von diesen Personen eingeschätzte Akzeptanz beziehungsweise soziale Validität (Schuler, 1990) der Test- und Fragebogenbatterie betrachtet.

8.2 Methoden

8.2.1 Stichprobe

Seit dem Jahr 2004 bis Mai 2006 wurde allen Personen, die das Self-Assessment komplett durchlaufen und die Rückmeldung gelesen hatten, eine Einladung zu einer Akzeptanzbefragung per Mail zugestellt. Von den in diesem Zuge angeschriebenen 8057 Personen haben 1960 den Fragebogen vollständig ausgefüllt. Somit beträgt die Rücklaufquote für diese Befragung in etwa 24 %. Das mittlere Alter beträgt in dieser Stichprobe 20,74 Jahre (Standardabweichung 3,968), der Median liegt bei 20 Jahren. 28,01 % der Teilnehmer sind weiblich, 71,99 % männlich.

8.2.2 Verwandte Testverfahren

In dieser Studie der von Kersting (2003b) entwickelte Fragebogen >Akzept!< zur Akzeptanz von Persönlichkeitsfragebogen und Leistungstests eingesetzt, ergänzt durch eine Skala, die für das Einsatzgebiet der Selbstberatung hier neu entwickelt wurde. Insgesamt umfasst der Akzeptanzfragebogen 25 Items. Die Teilnahme erfolgte komplett webbasiert und konnte in etwa 10 Minuten absolviert werden.

Im Allgemeinen sind Formulierungen von Akzeptanzfragebogen wie die hier eingesetzten von der jeweiligen Einsatzsituation und im besonderen Maße auch von der Art des verwandten Instruments, dessen Einsatz evaluiert werden soll, abhängig (Kersting, 1998). Aus diesem Grund werden im Originalfragebogen >Akzept!< unterschiedliche Itemformulierungen für die Akzeptanzbefragung im Rahmen von Persönlichkeitsfragebogen beziehungsweise im Rahmen von Leistungstests verwandt. Da das Self-Assessment sowohl aus Test- als auch aus Fragebogen-Bestandteilen besteht, wurden bei dieser Studie eingesetzten Akzeptanzfragebogen beide Formen des >Akzept!< vereint. Zudem sind alle Items – wie auch vom Autor (Kersting, 2003b) empfohlen – nicht nur an den Aufgabentyp sondern auch an das besondere Einsatzgebiet und

an die Zielgruppe angepasst: So wurden alle Itemformulierungen beibehalten, lediglich die Begriffe *Beruf*, *Berufsalltag* etc. wurden durch die jeweiligen Pendanten *Studium*, *Studienalltag* etc. ersetzt.

Insgesamt wurden 15 Items aus dem >Akzept!< verwandt, die sich auf folgende Skalen verteilen:

- Eine hohe *Augenscheinvalidität* ist nach Kersting (2003a) gegeben, „wenn die subjektiv wahrgenommene Messintention mit der seitens der Testperson vermuteten oder ihr explizit genannten diagnostischen Fragestellung übereinstimmt, zu deren Beantwortung das betreffende Verfahren eingesetzt wird“ (S. 54). Insgesamt vier Items bilden die entsprechende Fragebogenskala (Cronbachs- $\alpha = 0,634$, $n = 1\,960$). Zum Beispiel: „*Dass man mit solchen Fragen/Aufgaben wie denen des Verfahrens geeignete Personen für einen Studienplatz finden kann, ist zu bezweifeln*“.
- *Messqualität*: Dabei schätzt die Person ganz subjektiv anhand von vier Items ein, inwieweit sich bestehende Unterschiede in der Eignung von Personen mit dem Testverfahren präzise abbilden lassen (Cronbachs- $\alpha = 0,753$). Zum Beispiel: „*Die Auswertung des Verfahrens kann einen zutreffenden Eindruck von der Studieneignung eines Menschen vermitteln*“.
- *Kontrollierbarkeit*: Mit dieser drei Items umfassenden Skala soll ergründet werden, inwieweit den Testteilnehmern vor der Testung verständlich gemacht wurde, wie die gestellten Aufgaben zu bearbeiten sind. Dabei wird also unter anderem nach der Qualität der Instruktion gefragt (Cronbachs- $\alpha = 0,758$). Zum Beispiel ist ein Item wie folgt formuliert: „*Bei der Bearbeitung der Fragen/Aufgaben des Verfahrens wusste ich jederzeit, was ich tun musste*“.
- *Belastungsfreiheit* erfasst mit vier Items die Über- oder Unterforderung während der Testung (Cronbachs- $\alpha = 0,713$). So lautet zum Beispiel ein Item: „*Die Aufgaben waren überwiegend zu schwer für mich*“.

Für den hier beschriebenen Anwendungsfall wurde zusätzlich die Skala *Beratung* entwickelt, die den subjektiven Wert der geleisteten Beratung für den Teilnehmer abschätzen lassen soll (Cronbachs- $\alpha = 0,784$). Ein Beispiel dieser sieben Items umfassenden Skala ist Item 14: „*Das Feedback hat mich bestärkt in meiner Entscheidung für ein bestimmtes Studienfach*“.

Alle Items sind mit einer Bewertungsskala von 1 („*trifft nicht zu*“) bis 6 („*trifft genau zu*“) ohne verbalisierte Zwischenstufen versehen. Jeweils hohe Ausprägungen deuten auf positive Aussagen hin.

Zusätzlich werden noch drei Gesamteinschätzungen mithilfe einer Notenskala von *sehr gut* bis *ungenügend* erfragt: Je nach Item wird dabei der Fokus verstärkt auf inhaltliche Aspekte der Testbatterie („*Welche Schulnote würdest du dem bearbeiteten Beratungstest insgesamt geben?*“),

auf das Feedback („*Welche Schulnote würdest du dem Feedback zur Studienberatung geben?*“) oder auf technische Aspekte der Umsetzung („*Welche Schulnote würdest du der technischen Umsetzung des Tests und des Feedbacks geben?*“) gelegt.

8.2.3 Kontrolle des selektiven Dropouts

Ein zentrales Problem des hier verwandten freien, webbasierten Konzepts ist, dass zwar zur Teilnahme aufgerufen wurde, die letztendliche Entscheidung für oder gegen eine Teilnahme aber bei der Person verblieb. Entsprechend zu beachten ist daher ein potentieller selektiver Dropout, der bestünde, wenn die Wahrscheinlichkeit einer Teilnahme der einzelnen Personengruppen mit deren individueller Akzeptanzeinschätzung hinsichtlich des Verfahrens kovarierte. Denkbar wäre zum Beispiel, dass Personen, die das Testverfahren oder das Feedback nicht gut bewerten, sich erst gar nicht die Mühe machen, den Akzeptanzfragebogen auszufüllen. Anders als bei Testungen in überwachten Räumen ist diese Problematik in diesem Einsatzgebiet nicht vollends zu lösen, sie lässt sich aber – wie hier geschehen – teilweise kontrollieren: Dafür wurde auch im Hauptteil des Self-Assessments nach einer Bewertung des Gesamtverfahrens gebeten („*Welche Schulnote würdest du dem gesamten System geben?*“). Ausgehend von dieser Frage, die in einer differenzierteren Form auch noch Bestandteil des eigentlichen Akzeptanzfragebogens ist (siehe Abschnitt 8.2.2), lassen sich Mittelwertsunterschiede zwischen den Teilnehmern der Akzeptanzbefragung und deren Dropout-Gruppe kalkulieren. Stimmen die Mittelwerte überein beziehungsweise deutet die Effektstärke auf einen nicht bedeutsamen Unterschied hin, so lässt sich zumindest hinsichtlich der Gesamtbenotung die Vermutung eines selektiven Dropouts widerlegen.

8.2.4 Technische Realisierung

Besucherstatistik

Um den Erfolg eines webbasierten Angebotes wie dem hier beschriebenen Beratungsinstrument quantifizieren zu können, stehen als Datengrundlage die Logfiles des jeweiligen Webservers zur Verfügung. Mithilfe dieser Dateien werden alle Zugriffe auf eine Webseite protokolliert. Es lassen sich somit Aussagen über die Anzahl der Besuche und deren Dauer abschätzen. Allerdings stellt das verwandte Protokoll im Internet (*TCP-IP*) nur eingeschränkt Informationen zur Verfügung: Jede Anfrage eines Clients, also eines Besuchers, stellt eine eigenständige Aktion dar, die alle erst nachträglich wieder zum Verhalten eines Besuchers zusammengeführt werden müssen. Somit ist ein Tracking des Users nicht zweifelsfrei möglich.

Die Logfiles des Servers wurden mit der Software LIRE (Stichting LogReport Foundation, 2004) analysiert. Diese Software ermöglicht die Auswertung der Requests (also die Anzahl der abgerufenen Dateien auf dem Webserver), des dabei entstandenen Transfervolumens und der Anzahl der Besuche. Als Basis dieser Analysen wird die IP-Adresse verwendet, die automatisch durch die Netzanbieter an jeden Internetbenutzer verteilt wird und zu einem bestimmten Zeitpunkt eine eindeutige Zuordnung der Person ermöglicht. Viele Internetanbieter (Provider) verwenden aber dynamische IP-Adressen. Sie halten also nicht für jeden Kunden eine eigene IP-Adresse bereit, die dieser mit jeder Einwahl erneut erhält, sondern vergeben mit jeder Einwahl eines Kunden irgendeine Adresse aus einem Pool. So ist es wahrscheinlich, dass zwei unterschiedliche Personen zeitversetzt mit derselben IP-Adresse gesurft und das Self-Assessment besucht haben. Auf der anderen Seite ist es unwahrscheinlich, dass eine Person zweimal das Self-Assessment besucht und auch zweimal dieselbe IP-Adresse erhalten hat. Um dennoch die Anzahl unterschiedlicher Besuche abschätzen zu können, wird davon ausgegangen, dass ein Zugriff von derselben IP-Adresse aus nach einer Pause von 30 Minuten einen neuen Besuch darstellt. Daher lässt die im Ergebnisteil berichtete Zahl der Besuche nicht den Schluss zu, wie viele unterschiedliche Personen die Seite besucht haben. Vielmehr lässt sich davon ausgehen, dass manche Personen die Seite mehrfach besuchten, zum Beispiel in der Schule, zur Testteilnahme von zu Hause aus und später nochmals zur Wiedergabe des Feedbacks. Entsprechend tauchen diese Personen auch mehrfach in der Statistik auf.

Neben dieser dynamischen Vergabe von IP-Adressen verfälschen auch lokal gecachte Seiten und vom Internetprovider verwandte Proxy-Server die erstellten Logfiles und führen zu einer Unterschätzung der Zugriffszahlen.

Fragebogendarbietung

Seit Februar 2004 wurde vom Server automatisch nach jeder Teilnahme am Self-Assessment eine Mail an den Teilnehmer verschickt, die einen Aufruf zur Teilnahme an der Akzeptanzbefragung inklusive eines Verweises enthielt. Durch dieses Vorgehen war gewährleistet, dass die Personen bereits das Testfeedback erhalten und auch gelesen haben, bevor die Akzeptanzbefragung begonnen wurde. Nur so konnte vermieden werden, dass alleinig das Testverfahren und nicht das gesamte Beratungssystem samt der Rückmeldung bewertet wurde. Zudem war es so möglich, erst nach einer Pause mit der erneuten Befragung zu beginnen: Der Aufruf des Verweises in der E-Mail reichte dafür aus. Ein besonderer Anreiz für das Ausfüllen des Akzeptanzbogens wurde durch die Verlosung von Amazon-Büchergutscheinen, Software etc., die allesamt von Unternehmen gespendet wurden (siehe auch Abschnitt 5.3.1), unter den Teilnehmern der Befragung geschaffen. Alle Akzeptanzbögen wurden pseudonymisiert verwaltet, so dass zwar die Identität der Personen gewahrt blieb, dennoch aber ein Zusammenführen der

Testergebnisse und der Akzeptanzbewertungen möglich waren. Der Fragebogen wurde wie auch das Testverfahren selbst mit Adobe Flash durch den Autor dieser Arbeit erstellt.

8.3 Ergebnisse

8.3.1 Zugriffs- und Teilnehmerzahlen

In dem hier betrachteten Zeitraum von der erstmaligen Freischaltung am 1. August 2002 bis zum 30. April 2006 sind etwa 205 000 Besuche¹ auf der Webseite <http://www.assess.rwth-aachen.de> verzeichnet worden² (siehe Abbildung 8.1).

**Anzahl der Besuche und Testteilnahmen
in den Jahren 2002 bis 2006**

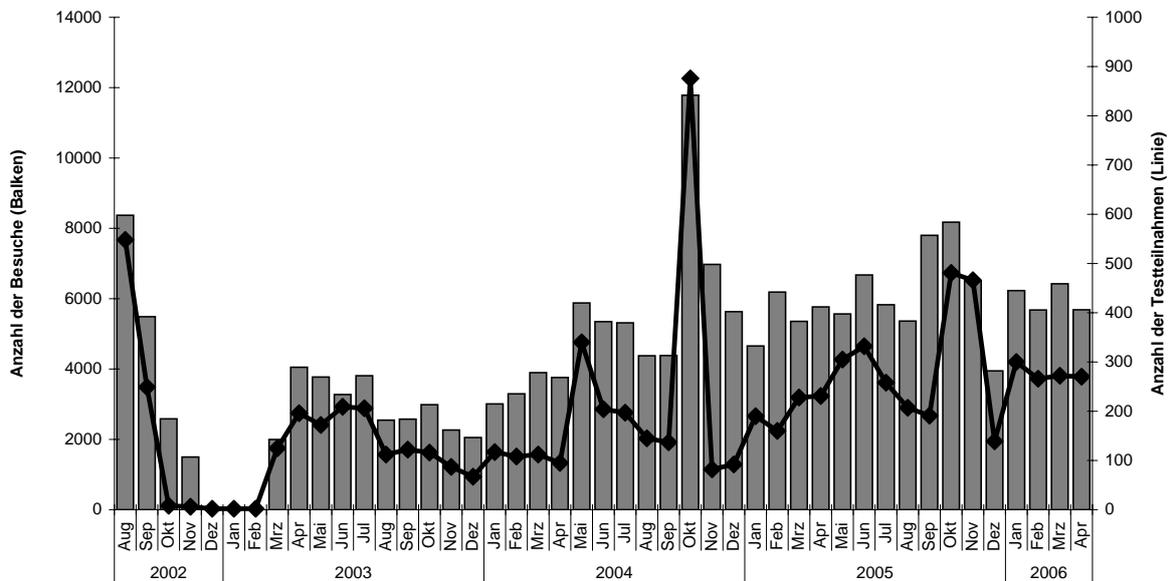


Abbildung 8.1: Anzahl der Seitenbesuche und der Testteilnahmen in den Jahren 2002 bis 2006¹

Im Mittel wurde die Webseite 165 Mal pro Tag konsultiert – mit steigender Tendenz über dem gesamten Zeitraum: So wurde 2002 das Beratungssystem durchschnittlich 154 Mal aufgerufen, zwei Jahre später bereits 175, 2005 197 und 2006 202 Mal pro Tag. Nur im Jahr 2003 wurden

¹ Wie bereits in Abschnitt 8.2.4 erwähnt, kann auf Basis der verwandten Analysetechnik nicht zweifelsfrei bestimmt werden, wie viele unterschiedliche Personen die Seite konsultiert haben. Daher wird hier nur die Anzahl der Besuche genannt. Rückschlüsse auf die Zahl der Besucher sind nicht – zumindest nicht zweifelsfrei – möglich. So werden zum Beispiel Personen mehrfach gezählt, die die Webseite nach einer längeren Zeit erneut besuchen.

² Zu dieser Analyse liegen aus technischen Gründen keine Logfiles aus dem Zeitraum zwischen dem 25. November 2002 und dem 9. März 2003 beziehungsweise zwischen dem 1. Mai und dem 11. Mai 2006 vor. In den Monaten November 2002 bis einschließlich Januar 2003 war das Self-Assessment zudem nicht für Testungen frei zugänglich.

aufgrund vergleichsweise geringer Marketinganstrengungen und einer längeren Offline-Zeit³ des Testsystems täglich etwa 54 Besuche weniger als im Vorjahr verzeichnet.

Bei einer näheren Betrachtung der Besuchszahlen fällt auf, dass die jeweiligen jährlichen arithmetischen Mittel der Besuche nur ein grobes Bild über die tägliche Anzahl der Besuche und deren Entwicklung über die Jahre ermöglichen: Bestimmte Peaks in den Besuchsanfragen beeinflussen die jährlichen Mittel enorm. Dieser Einfluss lässt sich auch an der vergleichsweise hohen Standardabweichung über alle Jahre von $SD = 115,136$ erkennen. Bei einer monatsweisen Analyse fallen insbesondere der August 2002, der Oktober 2004 und die Monate September und Oktober des Jahres 2005 auf (siehe auch Grafik 8.1). Im August 2002 – genauer am 23. August 2002 – wurde auf der Newsticker-Seite des IT-Zeitschriftenverlages Heise (die Nachricht lässt sich unter <http://www.heise.de/newsticker/result.xhtml?url=/newsticker/meldung/30194> abrufen) das Self-Assessment mit einer Meldung samt einem Hyperlink auf das System erwähnt. Alleine an diesem und dem folgenden Tag haben insgesamt 3 181 Besuche – oder etwa 18,0 % der Jahresbesuche 2002 – stattgefunden. In den Monaten September bis November der Folgejahre wurde in zahlreichen Marketingaktionen auf das Beratungssystem hingewiesen (siehe auch Abschnitt 5.3.1) – mit den entsprechenden Auswirkungen auf die Besucherzahlen. Zudem wurde in den Jahren 2004 und 2005 die Studie zur Kriteriumsvalidierung (siehe Abschnitt 7.4) durchgeführt. Alle Studienanfänger der Informatik (Haupt- und Nebenfach) der RWTH Aachen mussten – vor allem im Monat Oktober, teilweise auch im November – am Beratungssystem teilnehmen.

Erhebliche Schwankungen der Besuchsaktivitäten lassen sich auch im Tagesverlauf nachweisen. Die entsprechende Visualisierung dieser Ergebnisse bietet die Grafik 8.2. Dabei sind – anders als in der Abbildung 8.1 – an der Ordinate die Anzahl der Requests abgetragen, also die Anzahl der Seitenanfragen, die von den einzelnen Teilnehmern an den Webserver gestellt wurden, nicht die Anzahl der Besuche auf der Webseite⁴. Berücksichtigt wird auch dabei die komplette Online-Zeit des Self-Assessments ab dem August 2002 bis zum Ende der Logfile-Protokollierung am 30. April 2006. Bei der Betrachtung des Kurvenverlaufs fällt das Maximum in den Zugriffen am späten Nachmittag und frühen Abend auf. Aber auch in den Nachtstunden bis etwa zwei Uhr lassen sich bedeutsame Zugriffszahlen verzeichnen. Dies legt den Schluss nahe, dass die meisten Personen nicht während des Schulunterrichts auf die Projektseite zugegriffen haben, sondern in der Privatzeit.

³ Auch wenn diese Tage bei der Mittelwertbildung nicht berücksichtigt wurden, zeigt diese Offline-Zeit zahlreiche indirekte Auswirkungen auf die Bekanntheit der Webseite und somit auf das berechnete Mittel. Aufgrund der Offline-Zeit und des ausgebliebenen Marketings wurden Verweise auf die Projektseite gelöscht, Suchmaschinen-Rankings entsprechend angepasst und persönliche Weiterempfehlungen unterlassen.

⁴ Eine Auswertung der Besuchsstatistik im zeitlichen Tagesverlauf ist mithilfe der verwandten Software zur Logfileanalyse (siehe Abschnitt 8.2.4) nicht möglich.

Zugriffe im Tagesverlauf

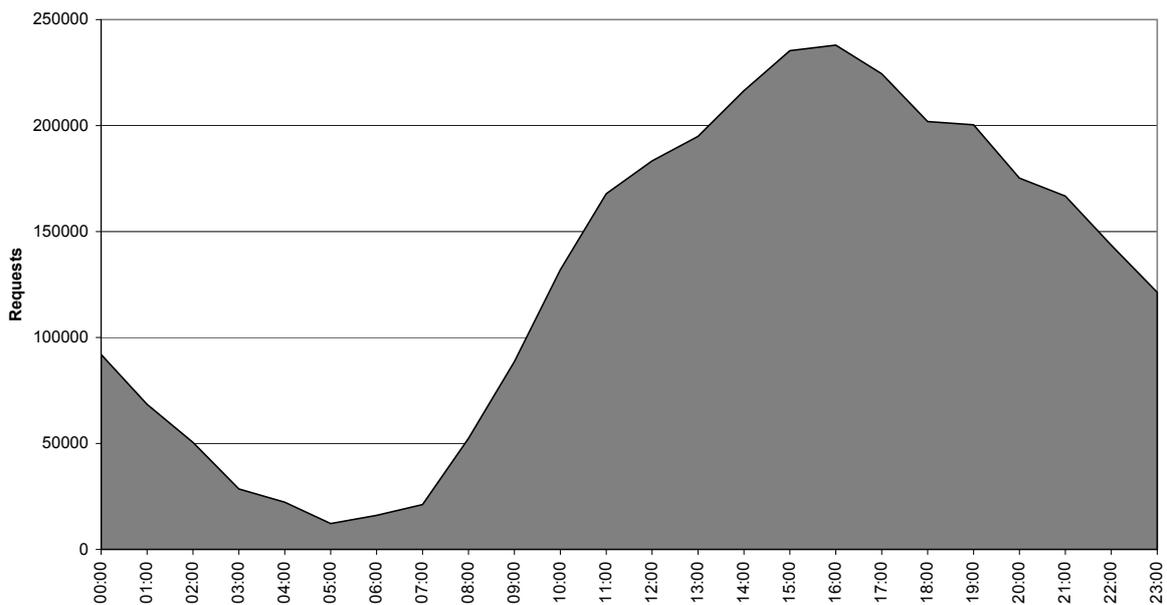


Abbildung 8.2: Anzahl der Seitenabfragen (Requests) im Tagesverlauf

Durch eine erweiterte Analyse der Logfiles bezüglich der übertragenen Referer⁵ lassen sich weitere Informationen über die Bekanntheit der Webseite ableiten. Ausgewertet wurden die Zugriffe in den letzten vier Monaten vor Abschluss der Datenerhebung. Wie erwartet sind mit einem Anteil von 26,0 % die meisten Besucher von der Internetseite der RWTH Aachen auf das Self-Assessment geleitet worden (<http://www-zhv.rwth-aachen.de>). Der zweitgrößte Anteil (20,7 %) der Besucher mit auslesbarem Referer hat Verweise vom Online-Lexikon Wikipedia (<http://de.wikipedia.org>)⁶ benutzt. Weitere Besuchergruppen mit Anteilen über 5,0 % sind von der Webseite des Magazins Stern (<http://www.stern.de>, 7,8 %), von der Seite der Universität Jena (<http://www.uni-jena.de>, 6,1 %) und von der deutschsprachigen Suchmaschinen-Seite Google (<http://www.google.de>, 5,5 %) verwiesen worden. Zusammen mit dem schweizerischen und dem österreichischen Ableger der Suchmaschine sind 6,0 % der Besucher auf Suchmaschinen-Treffer in Google zurückzuführen. Am häufigsten wurden die Suchbegriffe *Self-Assessment*, *Aachen*, *Informatik* und *RWTH* verwendet. Zahlreiche weitere Internetseiten (wie zum Beispiel <http://www.zeit.de>, <http://www.einstieg-informatik.de>, <http://www.zukunft-durch-technik.de>, <http://www.kompass-berufswahl.de> etc.) bilden zusammen den Rest der Anfragen, immerhin insgesamt noch 33,0 % der Zugriffe.

⁵ Sendet ein Client – also ein Webbrowser eines Teilnehmers – eine Seitenanfrage an einen Webserver, so werden zusätzlich Informationen über den Browsertyp, über das verwandte Betriebssystem etc. übertragen. Im Rahmen dieser Meta-Daten wird ein so genannter Referer übermittelt. Dieser umfasst die Information, von welcher Seite der Besucher durch Anklicken eines Internet-Verweises auf die abgerufene gelangt ist. Wird also zum Beispiel ein Besucher von der Internetpräsenz der Universität Jena auf die Self-Assessment-Projektseite weitergeleitet, so wird die Seite *www.uni-jena.de* im Referer vermerkt und kann dann vom Self-Assessment ausgelesen werden.

⁶ Informationen, in welchem Zeitraum die Links auf die Self-Assessment-Projektseite verwiesen haben, aber auch auf welcher Unterseite – zum Beispiel von Wikipedia – die Verweise erschienen sind, liegen nicht vor.

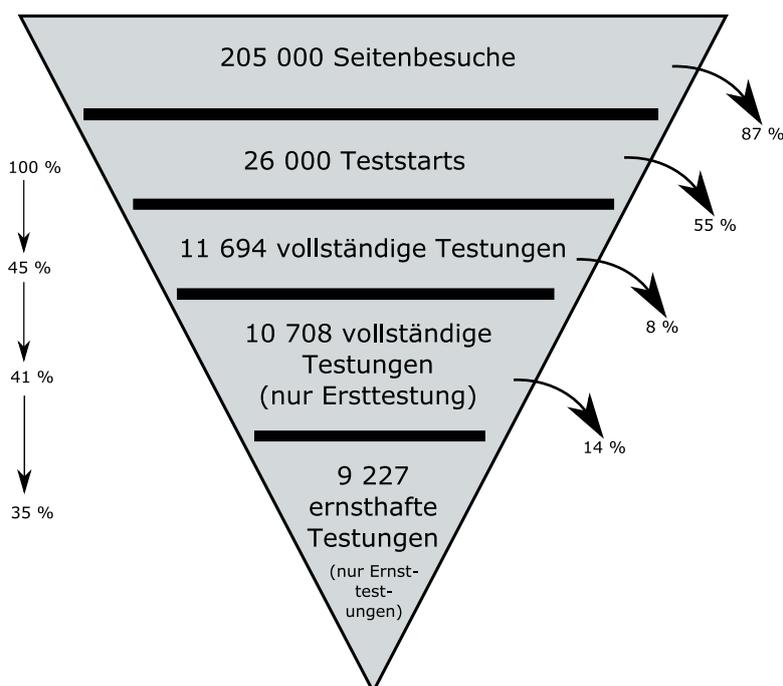


Abbildung 8.3: Überblick über die Anzahl der Besuche, Testungen und Testteilnehmer

Bis hierhin wurden lediglich die Anzahl der Seitenbesuche oder der Seitenabfragen ausgewertet. Relevanter erscheint eigentlich, wie viele Testungen in der hier betrachteten Zeit vollständig absolviert wurden. Das Zusammenspiel von Seitenbesuchen und Teilnehmerzahlen verdeutlicht die Abbildung 8.1. Auffällig sind die weitgehend vergleichbaren Verläufe. Entsprechend zeigt sich eine hohe Profilähnlichkeit der beiden Größen.

Bei den etwa 205 000 Besuche wurden ca. 26 000 Testungen gestartet (siehe Grafik 8.3). Von diesen Starts wurden insgesamt 11 694 Testungen abgeschlossen. Somit sind 55,0 % der gestarteten Testungen vor dem Absenden der Daten abgebrochen worden⁷. Da die Mehrfachteilnahme an dem Beratungssystem nicht ausgeschlossen war, eine solche aber im System vermerkt wurde, konnten diese Retests nachträglich eliminiert werden: So stammen etwa 8,0 % der Testungen von Teilnehmern, die bereits vorher einen Testdurchlauf komplett absolviert haben. Genau haben insgesamt 10 708 unterschiedliche Personen an einer Testung teilgenommen. 7 740 haben davon die letzte Testrevision dargeboten bekommen. Die bis hierhin genannten Zahlen berücksichtigen nicht, dass zahlreiche Personen das Testsystem zwar ausprobieren oder am Gewinnspiel teilnehmen wollten, aber nicht das komplette Verfahren ernsthaft durchlaufen haben. Bezüglich des hier verwandten Data-Screenings (siehe Abschnitt 5.3.1) reduziert sich die Anzahl der gesamten Testungen auf 9 227 (alle Tests bis auf *HK*) beziehungsweise auf 6 636 (nur *HK*). Somit haben etwa 35,0 % aller Teststarts auch zu verwertbaren Datensätzen geführt.

⁷ Seit Februar 2004 erfasst das Self-Assessment-System, wie viele Testungen gestartet und wie viele beendet wurden. Die dabei berechnete Testabbrucherquote von 55,0 % wurde hier für die gesamte Erhebungszeit als Berechnungsgrundlage verwandt.

8.3.2 Akzeptanzbefragung

Von den insgesamt 8 057 ausgegebenen Links zum Akzeptanzfragebogen wurden 1 960 Fragebogen vollständig ausgefüllt. Dies entspricht mit 24,0 % zwar der üblichen Rücklaufquote von webbasierten Umfragen (Batinic & Bosnjak, 2000; Trouteaud, 2004), gefährdet aber dennoch die Generalisierbarkeit der Ergebnisse (externe Validität). Wie bereits beschrieben (siehe Abschnitt 8.2.3) wird der Einfluss eines selektiven Dropouts⁸ kontrolliert, der sich in einer entsprechend geringen Rücklaufquote widerspiegeln könnte. Dabei steht im Mittelpunkt der folgenden Betrachtung, ob die Personen, die nicht an der Akzeptanzbefragung teilgenommen haben, wahrscheinlich das Verfahren grundsätzlich anders bewerten würden als die Teilnehmer der Befragung. Um dies näher analysieren zu können, wurde eine allgemeine Gesamtbewertung von allen Teilnehmern bereits vor dem Ende des Self-Assessments abgefragt⁹. Es lässt sich ein geringer Bewertungsunterschied zwischen den Personen nachweisen, die den Akzeptanzfragebogen ausfüllten ($M = 2,105$, $SD = 0,895$, $MD = 2,000$, $n = 1\,960$) und denen, die den Bogen nicht ausfüllten ($M = 2,171$, $SD = 0,858$, $MD = 2,000$, $n = 6\,097$). Aufgrund der hohen Stichprobengröße wird dieser Mittelwertvergleich (t-Test für unabhängige Stichproben) bei $\alpha = 0,050$ signifikant ($t = 2,9444$; $df = 8\,055$, $p = 0,003$). Die stichprobenunabhängige Effektgröße nach Cohen (1988) deutet aber mit einem Wert von $d = 0,080$ auf einen vernachlässigbaren Unterschied hin. Somit lässt sich zwar nicht allgemein ein selektiver Dropout hinsichtlich jeder Bestimmungsgröße ausschließen, bezüglich einer der wesentlichsten Größen, nämlich der Gesamtbewertung des Self-Assessments als Studienberatungssystem, lässt sich nicht ein solcher Effekt aber nachweisen. Aus diesem Grund werden im Folgenden ausschließlich die Ergebnisse der Analyse des umfangreicheren und daher aussagekräftigeren Akzeptanzbogens betrachtet.

69,5 % der Teilnehmer am Akzeptanzfragebogen ($n = 1\,960$) bewerten das Beratungssystem global im sechsstufigen Schulnotensystem mit den Noten *sehr gut* oder *gut*. Die mittlere Note beträgt dabei 2,320 mit einer Standardabweichung von 0,893. Verglichen mit der Gesamtbewertung vor der Rückmeldung, zeigt sich nach dieser eine um 0,2 Notenpunkte geringere Bewertung. Bei alleiniger Bewertung des Feedbacks zeigen sich leicht schwächere Akzeptanzwerte: Dabei geben 59,2 % der Personen mindestens die Note *gut*. Im Mittel wird die Qualität der Rückmeldung mit der Note 2,430 ($SD = 1,054$) bedacht. Die technische Umsetzung wird im Allgemein am besten beurteilt: 80,4 % der Personen vergeben mindestens eine gute Bewertung ($M = 1,851$ und $SD = 1,023$). Der Median aller drei globalen Bewertungen im Schulnotensystem liegt bei 2,000.

⁸ Denkbar wäre es zum Beispiel, dass nur Personen die freiwillige Rückmeldung abgeben, die das Self-Assessment insgesamt positiv bewerten.

⁹ Personen, die das Self-Assessment überhaupt nicht vollständig durchlaufen haben, werden hier nicht näher betrachtet. Daher wird unter Dropout hier verstanden, dass Personen das Self-Assessment zwar komplett durchlaufen und die Rückmeldung angezeigt bekommen, aber den Akzeptanzfragebogen nicht ausfüllen.

51,6 % der Teilnehmer schätzen ein, dass sie tendenziell mehr Informationen für die Studienfachentscheidung durch das Self-Assessment erhalten haben. 48,1 % der Personen haben nach der Teilnahme, primär durch die Rückmeldung und die Verweise auf weitere Webseiten, ein besseres Gefühl dafür, welche Anforderungen an der Hochschule auf sie warten. 53,5 % der Teilnehmer fühlen sich aufgrund des Beratungssystems in ihren vorigen Studienfachüberlegungen bestätigt. 23,8 % der Teilnehmer planen „auf jeden Fall“, Bekannten eine Teilnahme am Self-Assessment zu empfehlen. Ähnlich wollen – allerdings mehr oder weniger wahrscheinlich – insgesamt 73,4 % der Befragten verfahren.

Interessant ist ferner, die Akzeptanzbewertungen getrennt nach den vier verwandten Skalen aus dem Fragebogen >Akzept<! (Kersting, 2003b) genauer zu betrachten (siehe Abschnitt 8.2.2): Dabei werden die höchsten Bewertungen für die Skala Kontrollierbarkeit ($M = 4,930$, $SD = 1,247$)¹⁰ abgegeben, die unter anderem die Verständlichkeit der Instruktionen beschreibt. Etwas schwächer werden die Beratungsqualität ($M = 3,991$, $SD = 1,410$), die Augenscheinvalidität ($M = 3,698$ und $SD = 1,244$) und die Messqualität ($M = 3,604$ und $SD = 1,175$) eingeschätzt (siehe ebenfalls Tabelle 8.1).

Die Darstellung der Ergebnisse hat bisher den Umstand missachtet, dass die einzelnen Bewertungen der Akzeptanz zwischen den betrachteten Teilnehmern stark streuen. Bei einer differenzierten interindividuellen Auswertung offenbart sich, dass Akzeptanzurteile nicht nur mit der Qualität der Testbatterie und dem Testsetting zusammenhängen, sondern im besonderen Maße auch von der eigenen Leistung im Verfahren (Kersting, 1998): So korreliert das Ergebnis im Logiktest mit der Akzeptanz-Gesamtbenotung mit $r = -0,229$ ¹¹, mit dem Score im Mathetest mit $r = -0,144$ und im Matrizentest mit $r = -0,175$ (jeweils $n = 1960$). Hohe Korrelationen

¹⁰ Alle Items sind mit einer Bewertungsskala von 1 („trifft nicht zu“) bis 6 („trifft genau zu“) ohne verbalisierte Zwischenstufen versehen. Hohe Ausprägungen deuten auf positive Aussagen hin.

¹¹ Analog zu den statistischen Beziehungen zwischen Testscores und Schulnoten treten hier negative Korrelationen auf, da die Ratingskala zur Gesamtschätzung des Verfahrens einer Schulnote ähnelt. Gute (also niedrige) Akzeptanzbewertungen gehen dabei tendenziell mit besseren (also höheren) Testergebnissen einher.

Tabelle 8.1
Mittlere Itembeantwortungen und die interne Konsistenz der fünf Akzeptanzskalen

Skala	M	SD	Itemanzahl	Interne Konsistenz
Belastungsfreiheit	3,931	1,333	4	0,713
Kontrollierbarkeit	4,930	1,247	3	0,758
Augenscheinvalidität	3,698	1,244	4	0,634
Messqualität	3,604	1,175	4	0,753
Beratungsqualität	3,991	1,410	7	0,784

Anmerkungen: Die verwandte Skala ist sechsstufig mit den Polen *trifft nicht zu* (1) und *trifft genau zu* (6). Hohe Ausprägungen deuten auf positive Bewertungen hin ($n = 1960$)

Tabelle 8.2
 Interkorrelationen zwischen Akzeptanzscores und Testscores

Skala	Akzeptanz- benotung gesamt	Akzeptanz- benotung Feedback	Akzeptanz- benotung Technik	Belastungs- freiheit	Kontrollier- barkeit	Augenschein- validität	Mess- qualität	Beratungs- qualität
<i>LOG</i>	-0,229 **	-0,162 *	-0,148 *	0,288 **	0,375 ***	0,157 *	0,136 *	0,252 **
<i>MATH</i>	-0,144 *	-0,074	-0,091	0,305 ***	0,221 *	0,099 *	0,095	0,184 *
<i>MAT</i>	-0,175 *	-0,144 *	-0,077	0,272 **	0,292 **	0,145 *	0,155 *	0,229 **
<i>SW</i>	-0,126 *	-0,161 *	-0,109 *	0,246 **	0,227 **	0,086	0,113 *	0,226 **
<i>ELM</i>	-0,114 *	-0,090	-0,082	-0,001	0,066	0,107 *	0,106 *	0,136 *
<i>HOT</i>	-0,132 *	-0,174 *	-0,101 *	0,162 *	0,156 *	0,108 *	0,137 *	0,197 *
<i>HOM</i>	-0,013	-0,072	0,010	0,170 *	0,123 *	0,001	0,010	0,085
<i>HOP</i>	-0,034	-0,093	-0,029	0,078	0,047	0,043	0,058	0,090

Anmerkungen: $n = 1960$; Skalenbezeichnungen sind im Abkürzungsverzeichnis erläutert. Die ersten drei Akzeptanzbeurteilungen sind mithilfe einer Schulnotenskala erfasst worden. Entsprechend zeigen negative Korrelationen einen gleichförmigen Zusammenhang zwischen Akzeptanzbewertung und Leistungen im Self-Assessment, d.h. bessere Akzeptanzbewertungen gehen mit besseren Testergebnissen einher.

Auf eine Angabe der Signifikanz wird verzichtet (siehe Abschnitt 5.3.4). Stattdessen werden die Korrelationen in die Effektgröße d transformiert. Ein Asterisk deutet auf eine geringe Effektstärke hin ($0,2 \leq d < 0,5$), zwei Asterisken deuten auf eine mittlere ($0,5 \leq d < 0,8$) und drei auf eine große ($d \geq 0,8$).

zeigen sich ebenfalls mit den Skalen *Kontrollierbarkeit* und *Belastungen*: Dabei korreliert die Akzeptanzbewertung mit dem Logiktest mit $r = 0,375^{12}$ bzw. $r = 0,288$, mit dem Mathetest mit $r = 0,221$ beziehungsweise $r = 0,305$ und mit dem Matrizentest mit $r = 0,292$ beziehungsweise $r = 0,272$. Die Persönlichkeitsdimensionen zeigen etwas schwächere Beziehungen zur Akzeptanz: Mit der Gesamtnote korreliert die extrinsische Lernmotivation mit $r = -0,114$, die Selbstwirksamkeit mit $-0,126$ und die Handlungsorientierung bei Tätigkeitsausführung mit $-0,131$. Mit den Skalen *Kontrollierbarkeit* und *Belastung* zeigt insbesondere der Selbstwirksamkeitsfragebogen Zusammenhänge ($r = 0,227$ und $0,247$). Eine Übersicht aller Korrelationen zeigt Tabelle 8.2.

Im Rahmen einer multiplen Regression (berechnet mit Bootstrapping, zur Erläuterung siehe Abschnitt 7.3.4) zeigt die gesamte Testbatterie eine Varianzaufklärung in der Gesamtakzeptanzbewertung von im Mittel 7,9 % ($0,055 \leq R^2 \leq 0,103$; $0,236 \leq R \leq 0,321$; 95 %-KI). Betrachten wir die standardisierten Beta-Gewichte, so zeigen *LOG* und *MATH* unter den kognitiven Skalen Konfidenzintervalle, die den Nullwert nicht mit einschließen. Für *HOT* und *ELM* gilt unter den nicht-kognitiven Skalen dasselbe.

8.4 Diskussion

Für den Erfolg eines Studienberatungssystems, aber eben auch eines Marketinginstruments, das um die besten Schulabsolventen wirbt, ist die Akzeptanz in der Zielgruppe von entscheidender Bedeutung. Nur wenn die Studieninteressierten den Testergebnissen trauen, sich – zumindest teilweise – auch in diesen Resultaten wiederfinden und schlussendlich die Empfehlungen als schlüssig erachten, entscheiden sie unter anderem auf Basis eines Self-Assessments. Zudem empfehlen sie in diesem Fall das System weiter, wodurch neue Teilnehmer – auch ohne teure Werbemaßnahmen – gewonnen werden können.

Die berichteten überaus hohen Teilnehmerzahlen sprechen eine klare Sprache: Webbasierte Beratungssysteme treffen den Puls der Zeit. Studieninteressierte sind hoch motiviert, umfassende Informationen zu erhalten. Sie planen aktiver als häufig angenommen ihre Zukunft und suchen leistungsorientiert und pragmatisch nach essentiellen Informationen, um ihr Leben zu meistern (Hurrelmann & Albert, 2002).

Ein Großteil der Teilnehmer traut dem Test eine gute bis sehr gute Beratungsleistung zu. Sie haben das Gefühl, dass sie für die Entscheidungsfindung mehr Informationen zur Verfügung haben und somit die Studienfachwahl auch sicherer vollziehen können als ohne eine Teilnahme am Self-Assessment. Entsprechend empfehlen sie das Beratungssystem zu einem großen

¹² Gute Akzeptanzwerte gehen hier ebenfalls mit guten Testergebnissen einher.

Prozentsatz weiter. Gerade aufgrund der doch subjektiv langen Testdauer und der anstrengenden Bestandteile sind diese Ergebnisse nicht selbstverständlich. Allerdings zeigt sich da ein wichtiger Zusammenhang: Die wahrgenommene Belastung durch die Testteilnahme, aber auch die subjektive Beratungsqualität korrelieren bedeutsam mit den rückgemeldeten kognitiven Testscores. Personen, die bessere Ergebnisse erzielt und dies auch entsprechend zurückgemeldet bekommen haben, bewerteten das Self-Assessment weniger belastend. Zudem schätzen sie die Beratungsleistung höher ein als Personen mit schwächerer Testleistung. Scheinbar suchen Teilnehmer vermehrt nach bestätigenden Aussagen im Beratungssystem. Sollten sie aber schwächere Ergebnisse erzielt haben als erhofft oder als erwartet, attribuieren sie diese weniger auf sich, sondern verstärkt auf eine mangelnde Qualität des Tests. Somit beeinflusst eine bereits gegebene Rückmeldung durch das System auch die Beantwortung des Akzeptanzfragebogens. Sinnvoll erscheint es aufgrund dieser unerwünschten Beeinflussung, den Akzeptanzfragebogen aufzutrennen: Testbezogene Fragebogenbestandteile sollten direkt vor der Rückmeldung dargeboten werden. Eher rückmeldebezogene Fragebogen-Bestandteile müssen natürlich weiterhin nach der Rückmeldung erscheinen. Dabei kann allerdings weiterhin eine Beeinflussung nicht vermieden werden.

Diskutiert werden muss aber noch, ob die Ergebnisse der umfassenden Akzeptanzanalyse, die nur von einem Teil der Self-Assessment-Teilnehmer bearbeitet wurde, auf die gesamte Stichprobe der Getesteten übertragbar sind. Die Rücklaufquote der umfassenden Befragung liegt – wie bereits berichtet – mit 24,0 % in einem zu anderen Studien vergleichbaren Bereich (siehe zum Beispiel Batinic & Bosnjak, 2000; Trouteaud, 2004). Denkbar ist dennoch, dass gerade Personen den Bogen ausgefüllt haben, die eher zufrieden mit dem Beratungssystem waren. Somit wäre diese Akzeptanzbefragung besser ausgefallen, als es das Stimmungsbild wirklich vermittelt.

Kontrolliert wurde der selektive Dropout über eine Abfrage der Gesamtbewertung bereits vor Ende des Beratungssystems, so dass Vergleichswerte aller Personen zur Verfügung stehen. Sollten nur Personen den ausführlicheren Akzeptanzbogen ausgefüllt haben, die das Self-Assessment insgesamt besser bewertet haben als diejenigen, die den Bogen nicht mehr ausfüllten, so müssten dabei große Mittelwertsunterschiede auffallen. Die sehr geringe Effektstärke bei diesem Gruppenvergleich deutet jedoch auf keinen selektiven Dropout hin. Hinsichtlich der Gesamtbewertung unterscheiden sich somit diese Gruppen nicht. Weiterhin denkbar wäre jedoch, dass die Gruppen sich hinsichtlich anderer Eigenschaften unterscheiden. Eventuell haben Personen, die eine hohe Belastung während der Teilnahme erfahren haben, sich der Rückmeldung nicht mehr aussetzen wollen. Durch diesen selektiven Dropout wäre die Bewertung der Belastung positiver ausgefallen als sich hier zeigte.

Nicht weiter beachtet wurde ebenfalls, dass es eventuell noch einen weiteren möglichen selektiven Dropout gibt: Nur etwa 45,0 % der Teststarter haben den Beratungstest beendet. Einige dieser Testabbrecher haben sich allerdings zu einem späteren Zeitpunkt dem Test erneut ge-

stellt. Bereinigt man die Zahl der Testabbrucher entsprechend, so haben 62,0 % der Personen, die mindestens einmal den Test gestartet haben, auch mindestens einmal den Test abgeschlossen. Von den Abbrechern sind aufgrund technischer Beschränkungen der Self-Assessment-Plattform weder Test- noch Akzeptanzdaten vorhanden. Somit kann es – trotz der beschriebenen Kontrolle – einen selektiven Dropout bei der Akzeptanzbefragung gegeben haben: Personen, die den Test schon frühzeitig beendet haben, empfanden den Test wahrscheinlich wenig aussagekräftig, zu schwierig, zu leicht etc. Um auch diese Beweggründe des Abbruchs genauer kontrollieren zu können, müsste die technische Plattform des Self-Assessments Testergebnisse von Testabrechern speichern können. Zudem wäre es sinnvoll (aber wahrscheinlich schwer realisierbar), auch Abbrecher mit einer Kurzform des Akzeptanzbogens zu befragen, um die Hintergründe des Abbruchs genauer analysieren zu können.

Kapitel 9

Diskussion

In dieser Arbeit wurden zwei unterschiedliche Themen genauer betrachtet, die nun auch hier diskutiert werden sollen: Zuerst werden praxisrelevante Erfahrungen rund um webbasierte Self-Assessments dargestellt. Diese sollen helfen, selber Self-Assessments zu konstruieren und zur Unterstützung der Studienberatung einzuführen, aber auch das hier dargestellte Self-Assessment zu modifizieren und zu verbessern. Der zweite Abschnitt des Diskussionsteils betrachtet genauer die eignungsdiagnostischen Fragenstellungen und deren Implikationen. Diese Erörterungen sind nicht lediglich auf Studienberatungssysteme oder gar nur auf das hier vorgestellte Self-Assessment bezogen, sondern ermöglichen die Beantwortung auch allgemein relevanter Fragestellungen in der Studieneignungsforschung. Dabei ist zu bedenken, dass die Studie der Kriteriumsvalidierung – anders als die meisten sonstigen deutschsprachigen Studien dieser Art – auf einer Vollerhebung eines Anfängerjahrgangs beruht, also nicht durch Effekte der Selbstselektion etc. beeinflusst wurde.

9.1 Self-Assessments als Instrument zur Verbesserung der Studienberatung

Eine Hochschule verfolgt mit der Entwicklung und dem Einsatz eines computerisierten Studienberatungssystems wie mit dem hier vorgestellten Self-Assessment der RWTH Aachen ganz unterschiedliche Ziele. Natürlich nutzen Hochschulen alle Formen der Studienberatung auch als Werbung um zahlreiche, aber besonders auch um die besten Studieninteressierten. Ein Studienberatungssystem wird sich also auch immer daran messen müssen, wie häufig Ratsuchende das System benutzt haben. Zudem kann eben nur dann eine spürbare Verbesserung der Abbrecherquote durch ein solches System eintreten, wenn möglichst viele Personen vor der Studienaufnahme das Beratungssystem durchlaufen haben. Die vierte Studie hat genau diesen wichtigen Zuspruch der Zielgruppe näher betrachtet: Insgesamt über 200 000 Besuche wurden in den 3,5

Jahren auf der Webseite <http://www.assess.rwth-aachen.de> verzeichnet. Zur Bewertung dieser Zugriffszahl müssen die besonderen Rahmenbedingungen berücksichtigt werden: Für die Projekt-Webseite wurde keine kostenpflichtige Werbung gebucht, vielmehr wurden die Personen meist durch Empfehlungen von Bekannten oder aber auch von diversen Webseiten auf die Self-Assessment-Seite geleitet. Dementsprechend lassen sich die Zugriffszahlen natürlich nicht vergleichen mit hochfrequenten Seiten großer Unternehmen oder mit den Startseiten populärer Hochschulen etc.

Die trotz allem hohen Zugriffszahlen verdeutlichen aber: Anders als heute teilweise wahrgenommen wird, sind Schüler durchaus leistungsorientiert und strebsam, informieren sich, realisieren die Tragweite von Entscheidungen und erkunden auch Möglichkeiten, solche Entscheidungen abzusichern. Diese positive Einschätzung entspricht weitgehend auch den Ergebnissen der Shell-Studie 2002 (Hurrelmann & Albert, 2002). Das Internet ist derzeit für viele Jugendliche das wichtigste Medium für die Freizeitgestaltung, aber auch für die Suche nach einem passenden Beruf oder Studienfach (Heine & Willich, 2006). Internetseiten und entsprechend auch das Self-Assessment sind rund um die Uhr an 365 Tagen im Jahr „geöffnet“. Komplette selbstbestimmt kann jeder Interessent somit – ohne zeitliche Beschränkungen, ohne Terminabsprache und ohne weite Anreisen – das Self-Assessment zur persönlich passenden Zeit besuchen. Diese Flexibilität wurde auch eindrücklich genutzt: So wurde in der vierten Studie gezeigt, dass nur etwa 47,0 % der Zugriffe innerhalb der üblichen Geschäftszeiten von 9 bis 17 Uhr stattfanden¹. Zudem sind etwa ein Viertel aller Zugriffe an einem Samstag oder Sonntag erfolgt.

Bisher wurden nur die Zugriffszahlen auf das Beratungssystem kommentiert, nicht aber die viel relevantere Anzahl an kompletten Beratungsleistungen: Fast 12 000 Personen haben die Webseite des Self-Assessments in der hier betrachteten Zeit nicht nur besucht, sondern haben zusätzlich den kompletten Test absolviert. Bei einer mittleren Testdauer von 90 Minuten entspricht diese Anzahl einer Gesamt-Beratungsdauer von 18 000 Stunden. Anders als bei der üblichen *Face-to-Face*-Beratung bestehen die Kosten für ein Self-Assessment primär aus fixen Posten, die also nicht abhängig von der Teilnehmerzahl sind. Variable Kosten, so zum Beispiel die Kosten des Internetverkehrs (*Traffic-Kosten*), sind zwar vorhanden, aufgrund eines enormen Preisverfalls in diesem Bereich jedoch nahezu bedeutungslos. Versucht man, die Entwicklungskosten mit der Beratungsleistung gegeneinander aufzurechnen, erhält man Folgendes: Wenn die Entwicklungskosten eines Self-Assessments 50 000 € betragen, so hätte jede Einzel-Testung Kosten unter 5 € verursacht. Betrachtet werden muss nun aber noch, wie viel Geld der Einsatz eines Beratungssystems einsparen würde: Durchschnittlich würden etwa 35,0 %² von diesen Ratsuchenden das Studium abbrechen (in diesem Fall ca. 4 200 Personen). Bei einem mittleren

¹ Dieser Anteil bezieht sich sogar auf alle Wochentage, also genauso auf Samstag und Sonntag. Eine differenziertere Einschränkung auf Werktage war nicht möglich.

² Die Abbrecherquote in der Informatik liegt nach Heublein, Schmelzer und Sommer (2005) bei 38,0 %, die in der Elektrotechnik bei 33,0 %.

Studienabbruch nach 7,6 Semestern (siehe auch Abschnitt 1.4.4) und mittleren jährlichen Kosten von 5 500 € pro Student (siehe Heublein, Spangenberg & Sommer, 2003; Organisation für Wirtschaftliche Zusammenarbeit und Entwicklung, 2005) müssten – damit sich die Entwicklung amortisiert – in der bisherigen Laufzeit nur drei Personen weniger aufgrund des Ergebnisses im Self-Assessment ihr Studium abbrechen. Zu beachten ist, dass diese Berechnung nur die direkten volkswirtschaftlichen Kosten, nicht aber die individuellen Kosten eines Studienabbruchs berücksichtigt.

Bis hierhin sind nur Personen berücksichtigt worden, die die Projekt-Webseite besucht und einen Test komplett absolviert haben. Derzeit ist es aus technischen Gründen nicht möglich, Genaueres über Personen zu erfahren, die die Self-Assessment-Testbatterie zwar starten, diese aber nicht abschließen. Etwa 55,0 % aller Testdurchgänge wurden vorzeitig abgebrochen (siehe Abschnitt 8.3). Einige dieser Testabbrecher haben sich allerdings zu einem späteren Zeitpunkt dem Test erneut gestellt. Bereinigt man die Zahl der Testabbrecher entsprechend, so haben 62,0 % der Personen, die mindestens einmal den Test gestartet haben, auch mindestens einmal den Test abgeschlossen. Da gerade die große Gruppe der Testabbrecher einen interessanten Forschungsbereich darstellt, sollen die Rahmenbedingungen hier noch näher betrachtet werden: Derzeit wird das komplette Self-Assessment auf den heimischen Rechner geladen. Dieses Vorgehen hat einen entscheidenden Vorteil: Die Personen können das Beratungssystem selbst durchlaufen, ohne ständig mit dem Internet verbunden zu sein. Auch wenn dieser Vorteil in den letzten Jahren aufgrund der zahlreichen *Flatrate*-Angebote der Internetanbieter in Deutschland weniger wichtig wurde, kann dies gerade für den weltweiten Einsatz solcher Systeme entscheidend sein. Anders als in Deutschland kommen viele internationale Studierende aus Gebieten der Erde, die nicht über vergleichbar stabile und kostengünstige Internetverbindungen verfügen. Für dieses Einsatzgebiet ist es ein entscheidender Vorteil, wenn der Test ohne bestehende Internetverbindung bearbeitet und am Schluss zum Senden und Auswerten der Daten mit dem Server verbunden werden kann. Allerdings ist mit diesem Vorteil gleichzeitig der Nachteil verbunden, dass keine Zwischenergebnisse an den Server übertragen werden können. Somit sind von Testabbrechern überhaupt keine Informationen vorhanden. Eine Lösung, die beide Vorteile miteinander vereint, wäre folgende: Die Self-Assessment-Testbatterie würde genauso wie jetzt auch schon von der Projekt-Webseite heruntergeladen werden. Der Teilnehmer könnte danach die Internetverbindung trennen. Das System würde dann weiterhin mit den bekannten Einschränkungen arbeiten. Trennt der Teilnehmer aber seine Verbindung zum Internet nicht (zum Beispiel weil er über eine DSL-Verbindung verfügt), so würde auch eine Verbindung zum Self-Assessment-Server aufrecht bestehen bleiben. Entsprechend könnten kontinuierlich Zwischenergebnisse an den Server übermittelt werden.

Mit dieser, allerdings technisch aufwändigeren Lösung könnten sehr interessante Fragestellungen näher untersucht werden: So könnte geklärt werden, ob die Abbrüche eher test- oder eher zeitbezogen stattfinden. Denkbar wäre es zum Beispiel, dass Personen meist nach einer

bestimmten Zeit, zum Beispiel nach 60 Minuten, abbrechen (*zeitbezogener Abbruch*). Bei diesem Ergebnis wäre zu überlegen, wie die Effizienz der Verfahren erhöht und die Verfahren aufgetrennt werden könnten. Sollte man versuchen, in einer möglichst kurzen Zeit zu validen Ergebnissen zu kommen, wäre die Entwicklung adaptiver Testverfahren (Hornke, 1982) zu bedenken. Möglich ist es aber auch, dass Personen weniger zeitabhängig, sondern verstärkt bei der Bearbeitung bestimmter Testverfahren abbrechen (*testbezogener Abbruch*). So wird zum Beispiel die Bearbeitung der Matrizentests als besonders anstrengend erlebt. Wie auf ein solches Ergebnis sinnvoll reagiert werden kann, ist schwierig zu beantworten: In verpflichteten Eignungstests würden diese Tests an den Anfang der Testbatterie oder zumindest an einen Zeitpunkt direkt nach einer Pause verschoben werden. Im Falle freiwilliger Self-Assessments könnte das Verschieben an den Anfang der Testbatterie das Drop-Out-Verhalten sogar verstärken. Abhängig von der Kriteriumsvalidität dieses Testbestandteils sollte darüber nachgedacht werden, ob der belastende Test nicht aus der Batterie entfernt werden sollte.

Diese bisher diskutierten Fragestellungen gehen davon aus, dass alle Personen ein vergleichbares Abbruchverhalten zeigen. Sehr wahrscheinlich ist es zudem, dass es interindividuelle Unterschiede gibt, die auch direkt mit den bisherigen Testergebnissen und/oder mit besonderen Persönlichkeitseigenschaften einhergehen. Mit einer Analyse von Interaktionen aus Testbesonderheiten und Persönlichkeitseigenschaften der Teilnehmer lassen sich so neue und interessante Forschungsthemen erarbeiten.

Eine häufig diskutierte Frage, die thematisch direkt an die Drop-Out-Debatte anschließt, ist, ob Self-Assessments nicht lieber verbindlich eingeführt werden sollten. Ein Ergebnis einer solchen verpflichtenden Einführung ist nahe liegend: Die Drop-Out-Quote würde mit hoher Wahrscheinlichkeit sinken. Somit würden – zumindest auf dem ersten Blick – mehr Personen von dem Studienberatungssystem profitieren, weil sie es auf jeden Fall durchlaufen müssen. Fraglich ist allerdings, ob die Personen, die ohne eine Verpflichtung ein freiwilliges System abbrechen, in diesem Fall das Verfahren nicht nur einfach lustlos abschließen würden, ohne wirklich mit diesem ernsthaft gearbeitet zu haben. Damit würde weder der Hochschule, noch den Testkonstrukteuren, aber natürlich auch nicht den Teilnehmern geholfen werden. Studienberatung kann und sollte nur eine freiwillige Hilfestellung darstellen. Jede Teilnahme-Pflicht würde die positiven Effekte korrumpieren.

Die Ergebnisse, gerade zur Akzeptanz des Systems, zeigen deutlich: Mehr unterstützen könnte das Self-Assessment die Unschlüssigen, wenn das Beratungssystem nicht wie bisher nur auf drei Fächer, nämlich Informatik, Elektrotechnik und Technische Informatik, eingeschränkt wäre. Sinnvoll erscheint es, einen Routing-Test für zahlreiche Fachgebiete zu entwickeln, der vor die einzelnen fachspezifischen Self-Assessments geschaltet ist. Mittels einer Selbsteinschätzung bezüglich der Eignung und Neigung zu bestimmten Fächern, der Angabe verschiedener Studieninhalte und Berufsbilder, sollte nach und nach zu einzelnen Self-Assessments geleitet werden.

So könnte das erweiterte Self-Assessment nicht nur der Eignungsüberprüfung für ein bestimmtes Studienfach dienen, sondern eben auch dem Entscheidungsprozess vorher. Gerade das große Angebot unterschiedlichster Studienfächer könnte leichter durch die Unschlüssigen exploriert werden, da diese nach dem Routing-Test automatisch gefiltert würden. Auf einer interaktiven Tour ließe sich anschaulich von dem Studienfach berichten. Professoren, Studenten und Studienberater könnten dabei zu Wort kommen.

Im Rahmen des Self-Assessment an der RWTH Aachen wurde die Zentrale Studienberatung direkt mit dem webbasierten System verbunden, da durch eine Verknüpfung der test- und der personenbasierten Beratung die Vorteile beider Methoden kombiniert werden konnten. Die Studienberater waren über das Self-Assessment informiert und zudem auch in der Interpretation der Ergebnisse geschult worden. Je nach Studienfach empfahlen sie Studieninteressierten zusätzlich zu den weiteren Angeboten der Studienberatung auch einen Besuch auf der Projekt-Webseite, wenn diese noch nicht bekannt war. Auf der anderen Seite wurden die Ansprechpartner sowohl der zentralen Studienberatung als auch der Fachstudienberatung in der Rückmeldung des Self-Assessments erwähnt. Es wurde darauf hingewiesen, dass die Personen die Rückmeldung ausgedruckt zum Gespräch mitnehmen können und sollen, damit der Studienberater mehr objektive Informationen zur Beratung zur Verfügung stehen hat.

9.2 Eignungsdiagnostische Fragestellungen

Im Folgenden sollen die eignungsdiagnostischen Fragestellungen näher erörtert werden. Dabei werden sowohl spezifische Ergebnisse zu dem hier dargestellten Self-Assessment (zum Beispiel die Kriteriumsvalidität) als auch allgemein gültige Ergebnisse zur Studienfachwahl und zur Prognose von Studienerfolg zusammengefasst.

Die vielleicht wesentlichste Frage ist, wie gut fachspezifische Voraussagen auf der Basis einer kurzen Testbatterie von 60 bis 120 Minuten überhaupt funktionieren können – zumal die Testungen auch noch über das Internet stattgefunden haben. Zwar haben die meisten Befragten die Frage bejaht, durch das System mehr Informationen zur Studienfachwahl bekommen zu haben. Von dieser Befragung ist allerdings lediglich auf eine hohe Augenscheinvalidität zu schließen, nicht auf eine empirisch nachweisbare Kriteriumsvalidität.

Die zweite Studie hat zeigen können, dass die Testbatterie nicht nur die allgemeine Studierfähigkeit aufdeckt, sondern dass sie Interessenten unterschiedlicher Fachpräferenz unterscheiden kann. Somit zeigt sich, dass (a) Personen mit unterschiedlichen artikulierten Studienfachpräferenzen auch signifikant unterschiedliche Eignungen und Neigungen haben und (b) dass sich diese Präferenzen auch durch das Self-Assessment vorhersagen lassen. Erwartungsgemäß konnte gerade der Fragebogen des beruflichen Interesses (*INT*) zwischen den Präferenzgruppen

differenzieren. Das gemessene Interesse ist und bleibt somit ein wesentlicher Prädiktor für das artikulierte Interesse bezüglich des Studienfachs. Aber auch die anderen Verfahren, insbesondere die kognitiven Verfahren *LOG* und *MATH* haben Unterschiede zwischen Fachpräferenzen aufzeigen können. Interessant ist dabei, dass zumindest die Abiturdurchschnittsnote nahezu nichts mit der Studienpräferenz zu tun hat. Dieses Phänomen gilt aber wahrscheinlich nur, weil lediglich die Abiturgesamtnote betrachtet wurde, nicht aber Einzelfächer. In dem hier berichteten Fall wäre es sinnvoll gewesen, mehr unterschiedliche Schulfächer mit den jeweiligen Noten in die Analyse aufzunehmen. Wahrscheinlich ist, dass zwar nicht die Abiturabschlussnote für die Unterscheidung *Informatik* und *Gesellschaftswissenschaften* entscheidend ist, einzelne Fachnoten wie zum Beispiel die der Mathematik aber schon. Außerdem wurden nur zwei bestimmte Fächer ohne zentralen NC miteinander verglichen: Hätte man versucht, die Präferenz für Medizin versus Geisteswissenschaften vorhersagen zu wollen, wäre dies wahrscheinlich allein schon aufgrund der Zulassungsbeschränkungen auch vollständig mit der Abiturdurchschnittsnote gelungen.

Fraglich ist, wie dieses Ergebnis der zweiten Studie zu interpretieren ist: Die Unterschiede in den mittleren Testscores lassen vermuten, dass die Studienpräferenz auf der Basis einer eigenen Selbstselektion durch die Person, meist wahrscheinlich hinsichtlich eigener Eignungs- und Neigungsabschätzungen, erfolgt. Entsprechendes wird in der *gravitational theory* angenommen (Wilk, Desmarais & Sackett, 1995). Damit wird der Ansicht widersprochen, dass Personen eher nach extrinsischen Gründen (zum Beispiel nach dem Ansehen des Studienfachs etc.) entscheiden. Unklar ist jedoch die Kausalrichtung zwischen der gemessenen Interessenstruktur und der Eignung. Das Interesse könnte sowohl die Eignungskomponenten beeinflussen als auch vice-versa. Gerade für den Bereich der Studienpräferenz, also für den Altersbereich zwischen 16 und 21 Jahren, sind keine Studien bekannt, die dies näher beleuchten.

Eine vielleicht noch entscheidendere Frage besteht darin, ob der Studienerfolg von Personen, die sich für Informatik interessieren und ein entsprechendes Studium auch tatsächlich aufnehmen, prognostiziert werden kann. Zahlreiche Studien belegen, dass dies mit kognitiven Tests funktionieren kann (siehe Abschnitt 3.3). Fraglich war nur, ob dies auch durch ein webbasiertes und zudem nur 90 Minuten langes Beratungssystem funktioniert.

Diese Frage wurde in der dritten Studie beantwortet: Sowohl die kognitiven als auch die nicht-kognitiven Skalen des Self-Assessments korrelieren bedeutsam mit den beiden erhobenen Studienerfolgskriterien (gemittelte Prüfungsnoten im Studienfach Informatik und Anzahl bestandener Prüfungen). Die nicht-kognitiven Skalen sind zudem in der Lage, inkrementell Varianz über die Abiturnote und die kognitiven Tests hinaus aufzuklären. Der Leistungsaspekt (also die Studiennoten) wird besser durch die kognitiven Skalen aufgeklärt, die Anzahl bereits abgelegter Prüfungen allerdings durch die nicht-kognitiven und kognitiven gleichermaßen. Stets klären beide Prädiktorgruppen Varianz inkrementell über die Abiturdurchschnittsnote hinaus

auf. Die nachgewiesene inkrementelle Validität der Motivations- und Volitionsfragebogen unterstreicht die praktische Bedeutung für die Vorhersage von Studienerfolg. Deren Einsatz ist aber trotz dieser Ergebnisse in der Studierendenauswahl nicht empfehlenswert, da die hohe Verfälschungsgefahr (*cheating*) von Fragebogenergebnissen die Validität stark einschränken würde. Dieses Dilemma lässt sich aber leicht dadurch lösen, dass die Studierendenberatung durch Self-Assessments und die Studierendenauswahl stärker als bisher als eine Einheit verstanden werden. Beide Komponenten haben ihre Vorteile beziehungsweise Einsatzgebiete und somit auch ihre Berechtigung für den Hochschuleinsatz. Wie bisher können dann die Self-Assessments zur Beratung sowohl kognitive als auch nicht-kognitive Dimensionen abprüfen, Studierfähigkeitstests zur Auswahl später hingegen ausschließlich Fähigkeitsaspekte, diese dann aber gerichtsfest, messgenauer und valider als die kürzeren Self-Assessment-Batterien. In dieser Kombination lassen sich die Vorteile der Selbst- und Fremdselektion kombinieren und die Nachteile minimieren.

Was könnte noch an den Tests verbessert werden? Itemanalytisch ist gerade der Fragebogen zum beruflichen Interesse mit hohen internen Konsistenzen gelungen. Aber auch die anderen Skalen zeigen Gütekenntwerte, die denen veröffentlichter Fragebogen und Tests gleichen. Sinnvoll scheint es dennoch, die kognitiven Skalen einer erneuten Revision zu unterziehen. Wie bereits beim Matrizenest erfolgt, könnten auch die anderen beiden kognitiven Skalen mit mindestens zwei weiteren, möglichst inhaltshomogenen Items verlängert werden. So würde die Messgenauigkeit erhöht werden, was eventuell auch zu einer Steigerung der Validität führt. Eine Steigerung durch eine stärkere Homogenisierung der Skalen sollte nicht versucht werden, könnte dadurch doch gleichzeitig die Kriteriumsvalidität sinken (Lienert & Raatz, 1998). Fraglich ist ferner, ob der 30 Items umfassende Test zur Handlungskontrolle in dem Umfang in der Testbatterie verbleiben sollte. Nach den Ergebnissen zur Kriteriumsvalidierung würde der alleinige Einsatz der Handlungsorientierung nach Tätigkeitsinitiierung (*HOT*) zur Prognose des Studienerfolgs ausreichen. Geklärt werden muss, warum – trotz korrekter Programmierung und Auswertung – die Handlungskontrolle bei Planung ein negatives β -Gewicht in der linearen Regressionsgerade zur Prädiktion des Studienerfolgs zeigt.

Gerade für die Beantwortung der Frage, ob sich die Entwicklung eines Self-Assessment lohnt, weil signifikant weniger Personen ihr Studium abbrechen und mehr Personen mit guten Abschlussnoten die Universität verlassen, ist es noch zu früh. Auch wenn die hier berichteten Kriteriumsdaten den Schluss zulassen, dass es auch mit kurzen, webbasierten Systemen möglich ist, eine Vorhersage des Studienerfolgs vorzunehmen, ist es unklar, welche Auswirkungen eine solche Beratung auf die Studienfachwahl hat. Genau hierzu liegen nur vereinzelt Untersuchungen vor. Eine breit angelegte Studie, die untersucht, inwieweit Personen nicht kongruente Studienfachinformationen aufnehmen und die Fachwahl daraufhin anpassen, fehlt sowohl in Deutschland als auch in den USA. Selbst einzelne Studien, die diese Frage für einzelne Studienfächer an bestimmten Universitäten zu beantworten suchen, beantworten diese Fragestellung nur sehr eingeschränkt. Wesentlich für eine breite Einführung von Self-Assessments wird sein,

präzise die moderierenden Einflussfaktoren auf die Steuerungsfunktion zu erkunden. Welchen Einfluss haben zum Beispiel die Attraktivität des Studienfachs, die Verdienstmöglichkeiten, das Ansehen des späteren Berufs, die Studiengebühren oder die Persönlichkeit des Studieninteressierten (zum Beispiel Selbstwirksamkeit, Leistungsmotivation)? Anzunehmen ist, dass der Empfehlung eines Self-Assessments immer dann nachgekommen wird, wenn sich eben noch keine klare Präferenz für ein bestimmtes Fach gebildet hat. Ist dies einmal geschehen, wird es umso schwieriger, nicht kongruente Informationen so zu vermitteln, dass eine entsprechende Steuerfunktion ausgeübt wird. Diese Studie hier konnte aufzeigen, dass es sich lohnt, verstärkt in die Beantwortung dieser Fragen zu investieren.

Auch wenn die Prognosefähigkeit erfolgreich demonstriert wurde, sollte klar sein, dass ein etwa 90-minütiger Test keine 100 %-ige Entscheidungssicherheit geben kann. Ein Self-Assessment kann eben nur eine weitere Facette und nicht die alleinige Hilfestellung auf dem Weg zur Studienfachwahl sein; nicht mehr, aber auch nicht weniger. Wichtig ist nur, diese Einschränkung den Teilnehmern, aber auch den Fachbereichen zu vermitteln, die ein Self-Assessment einführen wollen.

Kapitel 10

Literaturverzeichnis

Abele, A. E., Andrä, M. S. & Schute, M. (1999). Wer hat nach dem Hochschulexamen schnell eine Stelle? Erste Ergebnisse der Erlanger Längsschnittstudie (BELA-E). *Zeitschrift für Arbeits- und Organisationspsychologie*, 43, 95-101.

Ach, N. (1935). *Analyse des Willens*. Berlin: Urban und Schwarzenberg.

Achtziger, A. & Gollwitzer, P. M. (2006). Motivation und Volition im Handlungsverlauf. In J. Heckhausen & H. Heckhausen (Hrsg.), *Motivation und Handeln* (S. 277-302). Heidelberg: Springer.

Ackerman, P. L. & Cianciolo, A. T. (2000). Cognitive, perceptual-speed, and psychomotor determinants of individual differences during skill acquisition. *Journal of Experimental Psychology: Applied*, 6, 259-290.

Ackerman, P. L. & Cianciolo, A. T. (2002). Ability and task constraint determinants of complex task performance. *Journal of Experimental Psychology: Applied*, 8, 194-208.

ACT Inc. (2005). *National Collegiate Retention and Persistence to Degree Rates*. Online in Internet: URL: http://www.act.org/path/policy/pdf/retain_2005.pdf (Stand 6. August 2006).

Althoff, K. (1986). Zur Aussagekraft von Schulzeugnissen im Rahmen der Eignungsdiagnostik. *Psychologie und Praxis*, 30, 77-85.

Amelang, M. (1978). Der Hochschulzugang. In K. J. Klauer (Hrsg.), *Handbuch Pädagogischer Diagnostik* (S. 1013-1022). Düsseldorf: Pädagogischer Verlag Schwann.

Amelang, M. (1997). Differentielle Aspekte der Hochschulzulassung: Probleme, Befunde, Lösungen. In T. Hermann (Hrsg.), *Hochschulentwicklung - Aufgaben und Chancen* (S. 88-105). Heidelberg: Asanger.

Amelang, M. & Zielinski, W. (1997). *Psychologische Diagnostik und Intervention* (2. Aufl.). Berlin: Springer.

- American Psychological Association (2001). *Publication manual of the American Psychological Association* (5. Aufl.). Washington, DC: American Psychological Association.
- Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A. (2001). *IST-2000R*. Göttingen: Hogrefe.
- Apenburg, E. (1979). Studienfachwahl und Studienmotivation. In B. Berendt, H. Gralki, H. Hecht & H. W. Hoefert (Hrsg.), *Hochschuldidaktik* (S. 19-22). Salzburg: O. Müller.
- Apenburg, E. (1980). Untersuchungen zur Studienzufriedenheit in der heutigen Massenuniversität. Frankfurt am Main: Lang.
- Artelt, C., Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Schümer, G., Stanat, P., Tillmann, K. & Weiß, M. (2001). *PISA 2000 - Zusammenfassung zentraler Befunde*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Assouline, M. & Meir, E. I. (1987). Metaanalysis of relationship between congruence and well-being measures. *Journal of Vocational Behavior*, 31, 319-332.
- Atkinson, R. C. (2002). Achievement versus aptitude in college admissions. *Issues in Science and Technology*, 18, 31-36.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2006). *Multivariate Analysenmethoden* (11. Aufl.). Berlin: Springer.
- Bahro, H. & Berlin, H. (2003). *Das Hochschulzulassungsrecht in der Bundesrepublik Deutschland: Kommentar* (4. Aufl.). Köln: Heymanns.
- Bandilla, W. (1999). WWW-Umfragen - Eine alternative Datenerhebungstechnik für die empirische Sozialforschung? In B. Batinic, A. Werner, L. Gräf & W. Bandilla (Hrsg.), *Online Research* (S. 9-19). Göttingen: Hogrefe.
- Bandura, A. (1977a). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191-215.
- Bandura, A. (1977b). *Social Learning Theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1983). Self-efficacy determinants of anticipated fears and calamities. *Journal of Personality and Social Psychology*, 45, 464-469.
- Bandura, A. (1997). *Self-efficacy*. New York, NY: Freeman.
- Barak, A. (1999). Psychological applications on the internet: A discipline on the threshold of a new millennium. *Applied & Preventive Psychology*, 8, 231-245.

-
- Barak, A. (2003). Ethical and professional issues in career assessment on the internet. *Journal of Career Assessment*, 11, 3-21.
- Bargel, T. (2005). *Soziale Herkunft der Studierenden im Vergleich von drei europäischen Regionen*. Online in Internet: URL: http://www.uni-konstanz.de/FuF/SozWiss/fg-soz/ag-hoc/News1/16-1Herkunft_drei-Regionen.pdf (Stand 6. Mai 2005).
- Baron-Boldt, J. (1989). *Die Validität von Schulabschlussnoten für die Prognose von Ausbildungs- und Studienerfolg*. Frankfurt am Main: Lang.
- Baron-Boldt, J., Schuler, H. & Funke, U. (1988). Prädiktive Validität von Schulabschlußnoten: Eine Metaanalyse. *Zeitschrift für Pädagogische Psychologie*, 2, 79-90.
- Barrick, M. R. & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1-26.
- Barrick, M. R., Mount, M. K. & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, 9, 9-29.
- Bartling, H. C. & Hood, A. B. (1981). An 11-year follow-up of measured interest and vocational choice. *Journal of Counseling Psychology*, 28, 27-35.
- Bartram, D. (2000). Internet recruitment and selection and assessment: Kissing frogs to find princes. *International Journal of Selection and Assessment*, 8, 261-274.
- Bartram, D. (2003). Testing through the internet. In R. Fernández-Ballesteros (Hrsg.), *Encyclopedia of Psychological Assessment* (S. 985-990). London: Sage Publications.
- Batinic, B. & Bosnjak, M. (2000). Fragebogenuntersuchung im Internet. In B. Batinic (Hrsg.), *Internet für Psychologen* (S. 287-317). Göttingen: Hogrefe.
- Baumert, J., Bos, W. & Lehmann, R. (2000). *Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe*. Opladen: Leske + Budrich.
- Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L. & Smith, T. A. (1996). *Mathematics achievement in the middle school years*. Chestnut Hill, MA: TIMSS International Study Center.
- Beauducel, A. & Kersting, M. (2002). Fluid and crystallized intelligence and the Berlin Model of Intelligence Structure (BIS). *European Journal of Psychological Assessment*, 18, 97-112.
- Becker, G. (1964). *Human capital*. New York, NY: Columbia University Press.
- Beil, C., Reisen, C. A., Zea, M. C. & Caplan, R. C. (1999). A longitudinal study of effects of academic and social integration and commitment on retention. *NASPA Journal*, 37, 376-385.

- Benbow, C. P. & Stanley, J. C. (1983). Sex difference in mathematical reasoning ability: More facts. *Science*, 222, 1029-1031.
- Bennett, R. E. (2005). *What does it mean to be a nonprofit educational measurement organization in the 21st century?* Princeton, NJ: Educational Testing Service.
- Bergmann, C. (1992). Schulisch-berufliche Interessen als Determinanten der Studien- bzw. Berufswahl und -bewältigung: Eine Überprüfung des Modells von Holland. In A. Krapp & M. Prenzel (Hrsg.), *Interesse, Lernen, Leistung. Neuere Ansätze der pädagogisch-psychologischen Interessenforschung* (S. 195-220). Münster: Aschendorff.
- Bergmann, C. (1994). Gemessene versus artikulierte Interessen als Prädiktoren der Berufs- bzw. Studienfachwahl und Anpassung im Studium. *Zeitschrift für Arbeits- und Organisationspsychologie*, 38, 142-151.
- Bergmann, C. (2003). Berufliche Interessentests - Wider die Anwendung geschlechtsspezifischer Normen. *Zeitschrift für Personalpsychologie*, 2, 66-77.
- Bergmann, C., Brandstätter, H. & Eder, F. (1994). *Studienberatungstests für Maturanten - Abschlußbericht*. Linz: Universität, Institut für Pädagogik und Psychologie.
- Bergmann, C. & Eder, F. (1991). Psychosoziale Umwelt als Bezugspunkt für Berufs- und Laufbahnentscheidungen. In H. Schuler & U. Funke (Hrsg.), *Eignungsdiagnostik in Forschung und Praxis. Beiträge zur Organisationspsychologie Band 10* (S. 294-296). Stuttgart: Verlag für Angewandte Psychologie.
- Bergmann, C. & Eder, F. (1992). *Allgemeiner Interessen-Struktur-Test (AIST)*. Weinheim: Beltz Test.
- Bergmann, C. & Eder, F. (2005). *Allgemeiner Interessen-Struktur-Test/Umwelt-Struktur-Test (AIST-R)*. Göttingen: Beltz Test.
- Betz, N. E. & Hackett, G. (1981). The relationship of career-related self-efficacy expectations to perceived career options in college women and men. *Journal of Counseling Psychology*, 28, 399-410.
- Birnbaum, M. H. (2000). *Psychological experiments on the Internet*. San Diego, CA: Academic Press.
- BITKOM (2006). *Hightech-Branche fordert Umdenken in der Zuwanderungspolitik*. Online in Internet: URL: http://www.bitkom.org/de/presse/8477_40427.aspx (Stand 26. Juli 2006).
- Black, A. E. & Deci, E. L. (2000). The effects of instructors' autonomy support and students' autonomous motivation on learning organic chemistry: A self-determination theory perspective. *Science Education*, 84, 740-756.

-
- Blickle, G. (1996). Personality traits, learning strategies, and performance. *European Journal of Personality*, 10, 337-352.
- Blum, F. (1980). *Studieneignung für die Ingenieurwissenschaften*. München: Minerva.
- Blum, F., Hensgen, A. & Trost, G. (1985). *Beratungstests für Oberstufenschüler und Abiturienten*. Bonn: Institut für Test- und Begabungsforschung.
- Boekarts, M. (1994). Action control: How relevant is it for classroom learning? In J. Kuhl & J. Beckmann (Hrsg.), *Volition and personality: Action- and state-oriented modes of control* (S. 427-435). Seattle, WA: Hogrefe & Huber.
- Borgen, F. H. & Seling, M. J. (1978). Expressed and inventoried interests revisited: Perspicacity in the person. *Journal of Counseling Psychology*, 25, 536-543.
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler* (5. Aufl.). Berlin: Springer.
- Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation für Sozialwissenschaftler* (2. Aufl.). Berlin: Springer.
- Bouffard-Bouchard, T. (1990). Influence of self-efficacy on performance in a cognitive task. *Journal of Social Psychology*, 130, 353-363.
- Brand (1987). The importance of general intelligence. In S. Modgil & C. Modgil (Hrsg.), *Arthur Jensen: Consensus and controversy* (S. 251-265).
- Brandstätter, E. (1999). Konfidenzintervalle als Alternative zu Signifikanztests. *Methods of Psychological Research Online*, 4, 1-17.
- Brandstätter, H. & Farthofer, A. (2003a). Einfluss von Erwerbstätigkeit auf den Studienerfolg. *Zeitschrift für Arbeits- und Organisationspsychologie*, 47, 134-145.
- Brandstätter, H. & Farthofer, A. (2003b). Erste Prüfungen - weiterer Studienerfolg. *Psychologie in Erziehung und Unterricht*, 50, 58-70.
- Brandstätter, H., Farthofer, A. & Grillich, L. (2001). Die Stabilität der Studienwahl als Funktion von Interessenkongruenz, Selbstkontrolle und intellektueller Leistungsfähigkeit. *Psychologie in Erziehung und Unterricht*, 48, 200-218.
- Brandstätter, H., Grillich, L. & Farthofer, A. (2002). Studienverlauf nach Studienberatung. *Zeitschrift für Pädagogische Psychologie*, 16, 15-28.
- Braun, L. G. (2001). *Wir brauchen eine neue Lernkultur!* Online in Internet: URL: <http://www.welt.de/daten/2001/12/06/1206fo300507.htx> (Stand 31. Januar 2005).

- Braxton, J. M., Sullivan, A. V. S. & Johnson, R. M. (1997). Appraising Tinto's theory of college student departure. In J. C. Smart (Hrsg.), *Higher education: Handbook of theory and research* (S. 107-164). New York, NY: Agathon Press.
- Bray, S. R. & Cowan, H. (2004). Proxy efficacy: Implications for self-efficacy and exercise intentions in cardiac rehabilitation. *Rehabilitation Psychology*, 49, 71-75.
- Bridgeman, B., McCamley-Jenkins, L. & Ervin, N. (2000). *Predictions of freshman grade-point average from the revised and recentered SAT I: Reasoning Test*. New York, NY: College Entrance Examination Board.
- Brocke, B. & Beauducel, A. (2001). Intelligenz als Konstrukt. In E. Stern & J. Guthke (Hrsg.), *Perspektiven der Intelligenzforschung* (S. 13-42). Lengerich: Pabst Science.
- Brown, S. & Gore, P. (1994). An evaluation of interest congruence indices: Distribution characteristics and measurement properties. *Journal of Vocational Behavior*, 45, 310-327.
- Brown, K. G., Le, H. & Schmidt, F. L. (2006). Specific aptitude theory revisited: Is there incremental validity for training performance? *International Journal of Selection and Assessment*, 14, 87-100.
- Brunstein, J. C. (1994). Dispositional action control as a predictor of how people cope with academic failure. In J. Kuhl & J. Beckmann (Hrsg.), *Volition and personality: Action- and state-oriented modes of control* (S. 341-350). Seattle, WA: Hogrefe & Huber.
- Buchanan, T. & Smith, J. L. (1999). Using the Internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology*, 90, 125-144.
- Bühl, A. & Zöfel, P. (2005). *SPSS 12 - Einführung in die moderne Datenanalyse unter Windows* (9. Aufl.). München: Pearson Studium.
- Bühner, M. (2004). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- Bundesinstitut für Bevölkerungsforschung (2004). *Bevölkerung - Fakten - Trends - Ursachen - Erwartungen* (2. Aufl.). Wiesbaden: Autor.
- Bundesministerium für Bildung und Forschung (2004a). Siebtes Gesetz zur Änderung des Hochschulrahmengesetzes. *Bundesgesetzblatt*, 47, 2298-2299.
- Bundesministerium für Bildung und Forschung (2004b). *Technologie und Qualifikation für neue Märkte*. Berlin: Autor.
- Bundesministerium für Bildung und Forschung (2005). *Hochschulrahmengesetz*. Online in Internet: URL: http://www.bmbf.de/pub/HRG_20050126.pdf (Stand 6. August 2006).

-
- Bundesministerium für Bildung und Forschung (2006). *Exzellenz in Bildung und Forschung - mehr Wachstum durch Innovationen*. Berlin: Autor.
- Bundesverband Digitale Wirtschaft (2006). *Langzeitstudie: Digitale Wirtschaft mit Voll-dampf voraus*. Online in Internet: URL: http://www.bvdw.org/ww/de/7_pub/aktuelles/pressemitteilungen.cfm?pID=771&fuseaction=showdetail (Stand 6. August 2006).
- Bundesvereinigung der Deutschen Arbeitgeberverbände (2003). *Arbeitgeber starten Offensive zur Nachvermittlung*. Online in Internet: URL: <http://www.bda-online.de/www/bdaonline.nsf/id/903DA7615A6AEBD3C1256DBA003C84E0> (Stand 6. August 2006).
- Burke, H. R. (1958). Raven's Progressive Matrices: A review and critical evaluation. *Journal of Genetic Psychology*, 93, 199-228.
- Burton, N. W. & Ramist, L. (2001). *Predicting success in college: SAT studies of classes graduating since 1980*. (Research Report No. 2001-2). New York City, NY: College Entrance Examination Board.
- Burton, N. W. & Wang, M. (2005). *Predicting long-term success in graduate school: A collaborative validity study (GRE Board Report No. 99-14R)*. Princeton, NJ: Educational Testing Service.
- Busato, V. V., Prins, F. J., Elshout, J. J. & Hamaker, C. (1999). The relation between learning styles, the Big Five personality traits and achievement motivation in higher education. *Personality and Individual Differences*, 26, 129-140.
- Busato, V. V., Prins, F. J., Elshout, J. J. & Hamaker, C. (2000). Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education. *Personality and Individual Differences*, 29, 1057-1068.
- Buttler, F. & Tessaring, M. (1993). Humankapital als Standortfaktor. *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung*, 26, 467-476.
- Cameron, J. & Pierce, W. D. (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational Research*, 64, 363-423.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press.
- Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, 40, 153-193.
- Cattell, R. B. (1965). *The scientific analysis of personality*. Harmondsworth: Penguin.
- Cattell, R. B. (1987). *Intelligence*. Amsterdam: Elsevier.

- Cervone, D. & Peake, P. K. (1986). Anchoring, efficacy, and action: The influence of judgmental heuristics on self-efficacy judgments and behavior. *Journal of Personality and Social Psychology*, 50, 492-501.
- Chemers, M. M., Hu, L. & Garcia, B. F. (2001). Academic self-efficacy and first year college student performance and adjustment. *Journal of Educational Psychology*, 93, 55-64.
- Christian, H. (1980). *Studierfähigkeit für das Fach Rechtswissenschaft*. München: Minerva.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Aufl.). Hillsdale, NJ: Erlbaum.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- Coyle, T. R. (2006). Test-retest changes on scholastic aptitude tests are not related to g. *Intelligence*, 34, 15-27.
- Cronbach, L. J. & Gleser, G. C. (1965). *Psychological tests and personnel decisions* (2. Aufl.). Urbana, IL: University of Illinois.
- Csikszentmihalyi, M. (1985). *Das flow-Erlebnis*. Stuttgart: Klett-Cotta.
- Dahl, G. (1971). Zur Berechnung des Schwierigkeitsindex bei quantitativ abgestufter Aufgabenbewertung. *Diagnostica*, 17, 139-142.
- Dawis, R. V. (2000). P-E fit as paradigm: Comment on Tinsley (2000). *Journal of Vocational Behavior*, 56, 180-183.
- Dawis, R. V. & Lofquist, L. H. (1984). *A psychological theory of work adjustment*. Minneapolis, MN: University of Minnesota Press.
- De Raad, B. (1996). Personality traits in learning and education. *European Journal of Personality*, 10, 185-200.
- De Raad, B. & Schouwenburg, H. C. (1996). Personality in learning and education: A review. *European Journal of Personality*, 10, 303-336.
- Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 18, 105-115.
- Deci, E. L., Koestner, R. & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125, 627-668.
- Deci, E. L. & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York City, NY: Plenum.

-
- Deci, E. L. & Ryan, R. M. (1993). Die Selbstbestimmungstheorie und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogische Psychologie*, 39, 223-238.
- Deidesheimer Kreis (1997). *Hochschulzulassung und Studieneignungstests*. Göttingen: Vandenhoeck & Ruprecht.
- Deutsche Presseagentur (2004). *Deutschland bei Pisa wieder unter Mittelmaß*. Online in Internet: URL: <http://www.faz.net/s/RubF7538E273FAA4006925CC36BB8AFE338/Doc~E93674E57D7C343DF9A7CD3F1BF1DFFBD~ATpl~Ecommon~Scontent.html> (Stand 8. April 2005).
- Deutsches Institut für Normung (2002). *DIN 33430 - Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen*. Berlin: Beuth.
- DeWitz, S. J. & Walsh, W. B. (2002). Self-efficacy and college student satisfaction. *Journal of Career Assessment*, 10, 315-326.
- Dezernat Planung Entwicklung und Controlling der RWTH Aachen (2002). *Zielvereinbarung zwischen dem Ministerium für Schule, Wissenschaft und Forschung des Landes Nordrhein-Westfalen und der Rheinisch-Westfälischen Technischen Hochschule Aachen*. Online in Internet: URL: http://www.rwth-aachen.de/zentral/pdf_zv_mswf_rwth.pdf (Stand 17. April 2004).
- Dickhäuser, O. & Plenter, I. (2005). "Letztes Halbjahr stand ich zwei" - Zur Akkuratheit selbst berichteter Noten. *Zeitschrift für Pädagogische Psychologie*, 19, 219-224.
- Die Zeit (2001). *Die Schule brennt*. Online in Internet: URL: http://www.zeit.de/archiv/2001/50/200150___+leiter.xml (Stand 8. April 2005).
- Die Zeit (2003). *Die Bilanz des Schreckens*. Online in Internet: URL: http://www.zeit.de/2003/50/B-Streitgespr_8ach (Stand 8. April 2005).
- Diehl, J. M. & Kohr, H. (1999). *Deskriptive Statistik* (12. Aufl.). Eschborn: Klotz.
- Diehl, J. M. & Staufenbiel, T. (2001). *Statistik mit SPSS Version 10.0*. Eschborn: Klotz.
- Diemand, A. & Schuler, H. (1991). Sozial erwünschtes Verhalten in eignungsdiagnostischen Situationen. In H. Schuler & U. Funke (Hrsg.), *Eignungsdiagnostik in Forschung und Praxis. Beiträge zur Organisationspsychologie Band 10* (S. 242-248). Stuttgart: Verlag für Angewandte Psychologie.
- Diseth, Å. (2003). Personality and approaches to learning as predictors of academic achievement. *European Journal of Personality*, 17, 143-155.
- Dixon, D. N. (2000). Frank Parsons. In A. E. Kazdin (Hrsg.), *Encyclopedia of psychology, Vol. 6*. (S. 63-64). New York City, NY: Oxford University Press.

- Dunn, D. S., McEntarffer, R. & Halonen, J. S. (2004). Empowering psychology students through self-assessment. In D. S. Dunn, C. M. Mehrotra & J. S. Halonen (Hrsg.), *Measuring up: Educational assessment challenges and practices for psychology* (S. 171-186). Washington, DC: American Psychological Association.
- Durkheim, E. (1973). *Der Selbstmord*. Neuwied am Rhein: Luchterhand.
- Ebel, R. L. (1979). *Essentials of educational measurement*. Englewood cliffs, NJ: Prentice-Hall.
- Edwards, J. R. (1991). Person-job fit: A conceptual integration, literature review, and methodological critique. In C. L. Cooper & I. T. Robertson (Hrsg.), *International review of industrial and organizational psychology, 1991, Vol. 6* (S. 283-357). Oxford: John Wiley & Sons.
- Efron, B. & Tibshirani, R. J. (1993). *An introduction to the bootstrap* (3. Aufl.). New York, NJ: Chapman and Hall.
- Eisenberger, R. & Cameron, J. (1996). Detrimental effects of reward: Reality or myth? *American Psychologist*, 51, 1153-1166.
- Elkins, S. A., Braxton, J. M. & James, G. W. (2000). Tinto's separation stage and its influence on first-semester college student persistence. *Research in Higher Education*, 41, 251-268.
- Engelhardt, W. H. (1993). Kommunikationspolitik als Ausgangspunkt des Wissenschaftsmarketing. In W. H. Engelhardt, J. König & T. Nietiedt (Hrsg.), *Wissenschaftsmarketing, Hochschule und Region im Umbruch* (S. 41-57). Bochum: Dr. N. Brockmeyer.
- Everson, H. T. (2004). Beyond individual differences: Exploring school effects on SAT scores. *Educational Psychologist*, 39, 157-172.
- Fleishman, E. A., Costanza, D. P. & Marshall-Mies, J. (1999). Abilities. In N. G. Peterson, M. D. Mumford, W. C. Borman, P. R. Jeanneret & E. A. Fleishman (Hrsg.), *An occupational information system for the 21st century: The development of O*NET* (S. 175-195). Washington, DC: American Psychological Association.
- Forschungsgruppe Wahlen Online (2006). *Internetstrukturdaten - Repräsentative Umfrage - 1. Quartal 2006*. Mannheim: Autor.
- Frackmann, E. (2005). *Warum sind die amerikanischen Spitzenuniversitäten so erfolgreich?* Hannover: Hochschul-Informations-System.
- French, J. W. (1951). *The description of aptitude and achievement tests in terms of rotated factors*. Chicago, IL: University of Chicago Press.

-
- Frey, M. C. & Detterman, D. K. (2004a). Scholastic assessment or g? The relationship between the scholastic assessment test and general cognitive ability. *Psychological Science*, 15, 373-378.
- Frey, M. C. & Detterman, D. K. (2004b). Scholastic assessment or g? The relationship between the scholastic assessment test and general cognitive ability: Erratum. *Psychological Science*, 15, 641.
- Fries, M. (2002). Abitur und Studienerfolg. Welchen Wert hat das Abitur für ein erfolgreiches Studium? *Beiträge zur Hochschulforschung*, 24, 30-51.
- Fuchs, J., Schnur, P. & Zika, G. (2005). Arbeitsmarktbilanz bis 2020. *IAB Kurzbericht*, 24, 1-4.
- Fuente, Á. d. I. & Ciccone, A. (2003). *Das Humankapital in der wissensbasierten globalen Wirtschaft*. Luxemburg: Amt für amtliche Veröffentlichungen der Europäischen Gemeinschaften.
- Geary, D. C. (1994). Children's mathematical development: Research and practical applications. In D. C. Geary (Hrsg.), *Individual differences in mathematical ability* (S. 131-153). Washington, DC: American Psychological Association.
- Georg, W. (2005). Die Reproduktion sozialer Ungleichheit im Lebenslauf. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, 2, 178-197.
- Germeijs, V. & Boeck, P. D. (2003). Career indecision: Three factors from decision theory. *Journal of Vocational Behavior*, 62, 11-25.
- Gesetzblatt für Baden-Württemberg (2005). Gesetz über die Hochschulen und Berufsakademien in Baden-Württemberg (Landeshochschulgesetz - LHG). *Gesetzblatt für Baden-Württemberg*, 1, 1-75.
- Gesk, I. (1999). *Studienabbruch an Pädagogischen Hochschulen – dargestellt am Studiengang für das Lehramt an Grund- und Hauptschulen*. Unveröffentlichte Dissertation, Ruprecht-Karls-Universität.
- Geyer, S. & Lilli, W. (1992). Handlungs- und Lageorientierung im Studium: Ein Vergleich zweier Fachrichtungen. *Zeitschrift für Sozialpsychologie*, 23, 119-128.
- Giesen, H., Böhmeke, W., Effler, M., Hummer, A., Jansen, R., Kötter, B., Krämer, H., Rabenstein, E. & Wener, R. R. (1981). *Vom Schüler zum Studenten*. München: Ernst Reinhardt.
- Giesen, H. & Gold, A. (1996). Individuelle Determinanten der Studiendauer. In J. Lompscher & H. Mandl (Hrsg.), *Lehr- und Lernprobleme im Studium* (S. 86-99). Bern: Hans Huber.

- Glaser, W. (1998). Induktion. In H. Häcker & K. H. Stapf (Hrsg.), *Dorsch Psychologisches Wörterbuch* (S. 394). Bern: Hans Huber.
- Gold, A. (1988). *Studienabbruch, Abbruchneigung und Studienerfolg*. Frankfurt am Main: Lang.
- Gold, A. & Kloft, C. (1991). Der Studienabbruch: Eine Analyse von Bedingungen und Begründungen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 23, 265-279.
- Gold, A. & Souvignier, E. (1997). Examensleistungen und Studiererleben bei Hochschulabsolventen. *Zeitschrift für Pädagogische Psychologie*, 11, 53-63.
- Gold, A. & Souvignier, E. (2005). Prognose der Studierfähigkeit - Ergebnisse aus Längsschnittanalysen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 37, 214-222.
- Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology*, 59, 1216-1229.
- Gollwitzer, P. (1996). Das Rubikonmodell der Handlungsphasen. In J. Kuhl & H. Heckhausen (Hrsg.), *Enzyklopädie der Psychologie: Motivation, Volition und Handlung (Themenbereich C, Serie IV, Band 4)* (S. 531-582). Göttingen: Hogrefe.
- Google (2006). *Google*. Online in Internet: URL: <http://www.google.de> (Stand 1. April 2006).
- Gore, P. A. (2006). Academic self-efficacy as a predictor of college outcomes: Two incremental validity studies. *Journal of Career Assessment*, 14, 92-115.
- Gosling, S. D., Vazire, S., Srivastava, S. & John, O. P. (2004). Should we trust web-based studies? *American Psychologist*, 59, 93-104.
- Gottfredson, G. D. (1999). John L. Holland's contributions to vocational psychology: A review and evaluation. *Journal of Vocational Behavior*, 55, 15-40.
- Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, 24, 79-132.
- Greening, T. (1998). Computer science: through the eyes of potential students. In ACM International Conference Proceeding Series (Hrsg.), *Proceedings of the third Australasian conference on computer science education* (S. 145-154). New York City, NY: ACM Press.
- Griesbach, H., Lewin, K., Heublein, U. & Sommer, D. (1998). *Studienabbruch - Typologie und Möglichkeiten der Abbruchquotenbestimmung*. Hannover: Hochschul-Informations-System.
- Grimm, J. & Grimm, W. (1854-1960). *Deutsches Wörterbuch*. Online in Internet: URL: <http://germazope.uni-trier.de/Projects/DWB> (Stand 2. Mai 2006).

-
- Guadagnoli, E. & Velicer, W. F. (1988). Relation to sample size to the stability of component patterns. *Psychological Bulletin*, 103, 265-275.
- Guilford, J., Christensen, P., Bond, N. & Sutton, M. (1954). A factor analysis study of human interests. *Psychological monographs: General and applied*, 68, 1-37.
- Häcker, H. & Stapf, K. (1998). Deduktion. In H. Häcker & K. Stapf (Hrsg.), *Dorsch Psychologisches Wörterbuch* (S. 168). Bern: Huber.
- Hänsgen, K. & Spicher, B. (2002). Numerus clausus: Finden wir mit dem "Eignungstest für das Medizinstudium" die Geeigneten? *Schweizerische Ärztezeitung*, 31, 1653-1660.
- Hannover, B. & Kessels, U. (2004). Self-to-prototype matching as a strategy for making academic choices. Why high school students do not like math and science. *Learning and Instruction*, 14, 51-67.
- Harackiewicz, J. M., Barron, K. E., Tauer, J. M. & Elliot, A. J. (2002). Predicting success in college: A longitudinal study of achievement goals and ability measures as predictors of interest and performance from freshman year through graduation. *Journal of Educational Psychology*, 94, 562-575.
- Harren, V. A. (1979). A model of career decision making for college students. *Journal of Vocational Behavior*, 14, 119-133.
- Heckhausen, H. (1989). *Motivation und Handeln* (2. Aufl.). Berlin: Springer.
- Heine, C., Bechmann, M. & Durrer, F. (2002). Wahrnehmung und Bedeutung der Arbeitsmarktaussichten bei Studienentscheidung und im Studienverlauf. In L. Bellmann & J. Velling (Hrsg.), *Arbeitsmärkte für Hochqualifizierte. Beiträge zur Arbeitsmarkt- und Berufsforschung Nr. 256* (S. 109-150). Nürnberg: Institut für Arbeitsmarkt und Berufsforschung.
- Heine, C., Briedis, K., Didi, H., Haase, K. & Trost, G. (2006). *Auswahl- und Eignungsfeststellungsverfahren beim Hochschulzugang in Deutschland und ausgewählten Ländern - HIS Kurzinformation A3/2006*. Hannover: Hochschul-Informationssystem.
- Heine, C., Scheller, P. & Willich, J. (2005). *Studienberechtigte 2005 - Studierbereitschaft, Berufsausbildung und Bedeutung der Hochschulreife*. Hannover: Hochschul-Informationssystem.
- Heine, C. & Willich, J. (2006). *Informationsverhalten und Entscheidungsfindung bei der Studien- und Ausbildungswahl*. Hannover: Hochschul-Informationssystem.
- Heldmann, W. (1984). *Studierfähigkeit*. Göttingen: Schwartz.
- Hell, B., Trapmann, S., Weigand, S., Hirn, J. O. & Schuler, H. (2005, September). *Die Validität von Prädiktoren des Studienerfolgs - eine Metaanalyse*. Vortrag auf der 4. Tagung der

- Fachgruppe Arbeits- und Organisationspsychologie der Deutschen Gesellschaft für Psychologie, Bonn.
- Hell, B. & Schuler, H. (2005). Verfahren der Studierendenauswahl aus Sicht der Bewerber. *Empirische Pädagogik*, 19, 361-376.
- Helmke, A. & Schrader, F. (2000). Procrastination im Studium - Erscheinungsformen und motivationale Bedingungen. In U. Schiefele & K. Wild (Hrsg.), *Interesse und Lernmotivation* (S. 206-226). Münster: Waxmann.
- Herkner, W. (1991). *Lehrbuch Sozialpsychologie* (5. Aufl.). Bern: Huber.
- Heublein, U., Schmelzer, R. & Sommer, D. (2005). *Studienabbruchstudie 2005*. Hannover: Hochschul-Informationssystem.
- Heublein, U., Schmelzer, R., Sommer, D. & Spangenberg, H. (2002). *Studienabbruchstudie 2002*. Hannover: Hochschul-Informationssystem.
- Heublein, U. & Sommer, D. (2002). *Studienanfänger 2000/2001 - Fachinteresse und berufliche Möglichkeiten bestimmen die Studienfachwahl*. Hannover: Hochschul-Informationssystem.
- Heublein, U., Sommer, D. & Weitz, B. (2004). *Studienverlauf im Ausländerstudium. Eine Untersuchung an vier ausgewählten Hochschulen*. Bonn: Deutscher Akademischer Austauschdienst.
- Heublein, U., Spangenberg, H. & Sommer, D. (2003). *Ursachen des Studienabbruchs: Analyse 2002*. Hannover: Hochschul-Informationssystem.
- Hinneberg, H. (2003). Abiturnote und Studienerfolg. *Hochschulwesen*, 51, 145-146.
- Hitpass, J. (1976). *Bericht über das Projekt Entwicklung von Anforderungsmerkmalen für die Studienfächer/-felder Biologie, Chemie, Physik; Bauingenieurwesen, Elektrotechnik, Maschinenbau; Mathematik; Rechtswissenschaften*. Köln: Pädagogische Hochschule Rheinland.
- Hochschulrektorenkonferenz (2006). *Der Hochschulkompass*. Online in Internet: URL: http://www.hochschulkompass.de/kompass/xml/index_stud.htm (Stand 12. März 2006).
- Höppel, D. & Moser, K. (1993). Die Prognostizierbarkeit von Studiennoten und Studiendauer durch Schulabschlussnoten. *Zeitschrift für Pädagogische Psychologie*, 7, 25-32.
- Hörner, W. (1999). Studienerfolgs- und Studienabbruchquoten im internationalen Vergleich. In M. Schröder-Gronostay & H. D. Daniel (Hrsg.), *Studienerfolg und Studienabbruch* (S. 1-15). Neuwied: Luchterhand.
- Höttecke, D. (2001). Die Vorstellungen von Schülern und Schülerinnen von der 'Natur der Naturwissenschaften'. *Zeitschrift für Didaktik der Naturwissenschaften*, 7, 7-23.

-
- Holland, J. L. (1997). *Making Vocational Choices* (3. Aufl.). Odessa, FL: Psychological Assessment Resources.
- Holling, H., Lüken, K., Preckel, F. & Stotz, M. (2000). *Berufliche Entscheidungsfindung: Bestandsaufnahme, Evaluation und Neuentwicklung computergestützter Verfahren zur Selbsteinschätzung*. Nürnberg: Zentralamt der Bundesanstalt für Arbeit.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.
- Horn, J. L. & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, 57, 253-270.
- Hornke, L. F. (1982). Testdiagnostische Untersuchungsstrategien. In K. J. Groffmann & L. Michel (Hrsg.), *Enzyklopädie der Psychologie, Band 2, Grundlagen psychologischer Diagnostik* (S. 130-172). Göttingen: Hogrefe.
- Hornke, L. F., Küppers, A. & Etzel, S. (2000). Konstruktion und Evaluation eines adaptiven Matrizen-tests. *Diagnostica*, 46, 182-188.
- Huber, H. P. (1973). Verallgemeinerungen des Cattellschen Profilähnlichkeitskoeffizient r_p unter dem Aspekt der klassischen Reliabilitätstheorie. *Zeitschrift für experimentelle und angewandte Psychologie*, 10, 39-53.
- Hubert, L. & Arabie, P. (1987). Evaluating order hypotheses within proximity matrices. *Psychological Bulletin*, 102, 172-178.
- Hubin, D. R. (1988). *The Scholastic Aptitude Test: Its development and introduction*. Online in Internet: URL: <http://darkwing.uoregon.edu/~hubin/> (Stand 25. Januar 2007).
- Hünninghausen, L. & Brümmer, V. (2002). *Die Besten gehen ins Netz: Report E-Recruitment: innovative Wege bei der Personalauswahl*. Düsseldorf: Symposion.
- Huldi, M. (1997). *Die Intelligenzstruktur von Mittelschülern*. Bern: Lang.
- Humphreys, L. G. & Swets, J. A. (1991). Comparison of predictive validities measured with biserial correlation and ROC's of signal detection theory. *Journal of Applied Psychology*, 76, 316-321.
- Hunter, J. E. & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Hurrelmann, K. & Albert, M. (2002). *Jugend 2002 - 14. Shell Jugendstudie*. Frankfurt am Main: Fischer-Taschenbuch-Verlag.

- Hurtz, G. M. & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology*, 85, 869-879.
- Ilgen, M., McKellar, J. & Tiet, Q. (2005). Abstinence self-efficacy and abstinence 1 year after substance use disorder treatment. *Journal of Consulting and Clinical Psychology*, 73, 1175-1180.
- Ingenkamp, K. (1991). *Die Bedeutung von Schultests für moderne Bildungssysteme*. Weinheim: Beltz.
- Institut der deutschen Wirtschaft Köln (2005a). *Den Fortschritt neu entdeckt*. Online in Internet: URL: <http://www.iwkoeln.de/default.aspx?p=pub&i=1665&pn=2&n=n1665&m=pub&f=4&a=18387> (Stand 6. August 2006).
- Institut der deutschen Wirtschaft Köln (2005b). *Mittelstand gerät in Zugzwang*. Online in Internet: URL: <http://www.iwkoeln.de/default.aspx?p=pub&i=1663&pn=6&n=n1663&m=pub&f=4&a=18381> (Stand 26. Juli 2006).
- Isserstedt, W., Middendorff, E., Weber, S., Schnitzer, K. & Wolter, A. (2004). *Die wirtschaftliche und soziale Lage der Studierenden in der Bundesrepublik Deutschland 2003*. Bonn: Bundesministerium für Bildung und Forschung.
- Jäger, A. O. (1973). *Dimensionen der Intelligenz* (3. Aufl.). Göttingen: Hogrefe.
- Jäger, A. & Althoff, K. (1983). *Der Wilde-Intelligenz-Test (WIT): Ein Strukturdiagnostikum*. Göttingen: Hogrefe.
- Jensen, A. R. (1998). *The g factor - the science of mental ability*. Westport, CT: Praeger.
- Jerusalem, M. & Schwarzer, R. (1999). *Allgemeine Selbstwirksamkeit*. Online in Internet: URL: http://web.fu-berlin.de/gesund/skalen/Allgemeine_Selbstwirksamkeit/hauptteil_allgemeine_selbstwirksamkeit.htm (Stand 28. Mai 2006).
- Jex, S. M., Bliese, P. D., Buzzell, S. & Primeau, J. (2001). The impact of self-efficacy on stressor-strain relations: Coping style as an explanatory mechanism. *Journal of Applied Psychology*, 86, 401-409.
- Jörin, S., Stoll, F., Bergmann, C., Eder, F. & Bundesagentur für Arbeit, (2004). *Explorix*. Bern: Huber.
- Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality*, 39, 103-129.
- Johnson-Laird, P. N. & Byrne, R. M. J. (1991). *Deduction*. Hove: Lawrence Erlbaum.

-
- Joinson, A. (1999). Social desirability, anonymity, and Internet-based questionnaires. *Behavior Research Methods, Instruments, & Computers*, 31, 433-438.
- Judge, T. A. & Bono, J. E. (2001). Relationship of core self-evaluations traits - self-esteem, generalized self-efficacy, locus of control, and emotional stability - with job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology*, 86, 80-92.
- Judge, T. A., Heller, D. & Mount, M. K. (2002). Five-Factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology*, 87, 530-541.
- Judge, T. A., Higgins, C. A., Thoresen, C. J. & Barrick, M. R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel Psychology*, 52, 621-652.
- Kahn, J. H. & Nauta, M. M. (2001). Social-cognitive predictors of first-year college persistence: The importance of proximal assessment. *Research in Higher Education*, 42, 633-652.
- Kazemzadeh, F., Minks, K. & Nigmann, R. (1987). 'Studierfähigkeit' - eine Untersuchung des Übergangs vom Gymnasium zur Universität. Hannover: Hochschul-Informationssystem.
- Kehm, B. M. (2004). Hochschulen in Deutschland. *Aus Politik und Zeitgeschichte*, 25, 6-17.
- Kersting, M. (1998). Differentielle Aspekte der sozialen Akzeptanz von Intelligenztests und Problemlösenszenarien als Personalauswahlverfahren. *Zeitschrift für Arbeits- und Organisationspsychologie*, 42, 61-75.
- Kersting, M. (1999). Personalauswahl mit dem BIS-r-DGP Test: Evaluation der prognostischen Validität. *DGP Informationen*, 47, 43-60.
- Kersting, M. (2003a). Augenscheinvalidität. In K. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 54-55). Weinheim: Beltz.
- Kersting, M. (2003b). Der >Akzept!< Fragebogen zur Messung der Akzeptanz diagnostischer Verfahren. *Unveröffentlichtes Testmanual*. Aachen: RWTH Aachen.
- Kersting, M. (2005, September). *Mythen, Vorurteile und Tabus rund ums Thema Auswahl und Beratung von Studienbewerber(innen)n*. Vortrag auf der 8. Arbeitstagung Differentielle Psychologie, Persönlichkeitspsychologie und Psychologische Diagnostik, Marburg.
- Kleine, D. & Jäger, A. O. (1989). Kriteriumsvalidität eines neuartigen Tests zum Berliner Intelligenzstrukturmodell - Eine Untersuchung an brasilianischen Schülern und Studenten. *Diagnostica*, 35, 17-37.

- Kobrin, J. L., Camara, W. J. & Milewski, G. B. (2002). *The Utility of the SAT I and SAT II for Admissions Decisions in California and the Nation*. College Board Research Report No. 2002-6. New York City, NY: College Entrance Examination Board.
- Koch, J. & Mohr, J. (2006, 11. Dezember). Gute Fächer, schlechte Fächer. *Der Spiegel*, S. 64-79.
- Konegen-Grenier, C. (2001). *Studierfähigkeit und Hochschulzugang*. Köln: Deutscher Instituts-Verlag.
- Kotter, J. P. (1978). *Self-assessment and career development*. Englewood Cliffs, NJ: Prentice-Hall.
- Krapp, A. (1992). Das Interessenkonstrukt - Bestimmungsmerkmale der Interessenhandlung und des individuellen Interesses aus der Sicht einer Person-Gegenstands-Konzeption. In A. Krapp & M. Prenzel (Hrsg.), *Interesse, Lernen, Leistung. Neuere Ansätze der pädagogisch-psychologischen Interessenforschung* (S. 297-329). Münster: Aschendorff.
- Krapp, A. (2003). Die Bedeutung der Lernmotivation für die Optimierung des schulischen Bildungssystems. *Politische Studien*, 54, 91-105.
- Krapp, A. & Ryan, R. M. (2002). Selbstwirksamkeit und Lernmotivation. *Zeitschrift für Pädagogik*, 44, 54-82.
- Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J. & Couper, M. (2004). Psychological research online - Report of board of scientific affairs' advisory group on the conduct of research on the internet. *American Psychologist*, 59, 105-117.
- Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukophadhyay, T. & Scherlis, W. (1998). Internet paradox: A social technology that reduces social involvement and psychological well-being? *American Psychologist*, 53, 1017-1031.
- Kristof, A. L. (1996). Person-organization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel Psychology*, 49, 1-49.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology*, 5, 213-236.
- Krosnick, J. A., Holbrook, A. L., Berent, M. K., Carson, R. T., Hanemann, W. M., Kopp, R. J., Mitchell, R. C., Presser, S., Ruud, P. A., Smith, V. K., Moody, W. R., Green, M. C. & Conaway, M. (2002). The impact of no opinion response options on data quality: Non-attitude reduction or an invitation to satisfice? *Public Opinion Quarterly*, 66, 371-403.
- Kuhl, J. (1983). *Motivation, Konflikt und Handlungskontrolle*. Berlin: Springer.

Kuhl, J. (1987). Motivation und Handlungskontrolle: Ohne guten Willen geht nichts. In H. Heckhausen, P. Gollwitzer & F. E. Weinert (Hrsg.), *Jenseits des Rubikon: Der Wille in den Humanwissenschaften* (S. 101-120). Berlin: Springer.

Kuhl, J. (o.J.). *Kurzanweisung zum Fragebogen HAKEMP 90*. Online in Internet: URL: <http://www.psych.uni-osnabrueck.de/fach/persoel/preprints/hakemp90.pdf> (Stand 28. Mai 2006).

Kultusministerkonferenz der Bundesrepublik Deutschland (2000). *Vereinbarung zur Gestaltung der gymnasialen Oberstufe in der Sekundarstufe II*. Online in Internet: URL: <http://www.kmk.org/doc/beschl/abi-01.pdf> (Stand 6. August 2006).

Kultusministerkonferenz der Bundesrepublik Deutschland (2003). *Prognose der Studienanfänger, Studierenden und Hochschulabsolventen bis 2020*. Online in Internet: URL: <http://www.kmk.org/statist/dok167.pdf.zip> (Stand 6. August 2006).

Kuncel, N. R., Crede, M. & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research*, 75, 63-82.

Kuncel, N. R., Credé, M., Thomas, L. L., Klieger, D. M., Seiler, S. N. & Woo, S. E. (2005). A meta-analysis of the validity of the pharmacy college admission test (PCAT) and grade predictors of pharmacy student performance. *American Journal of Pharmaceutical Education*, 59, 339-347.

Kuncel, N. R., Hezlett, S. A. & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance. *Psychological Bulletin*, 127, 162-181.

Kuncel, N. R., Hezlett, S. A. & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, 86, 148-161.

Levesque, R. (2005). *SPSS Programming and Data Management* (2. Aufl.). Chicago, IL: SPSS Inc.

Lewin, K. (1999). Studienabbruch in Deutschland. In M. Schröder-Gronostay & H. Daniel (Hrsg.), *Studienerfolg und Studienabbruch: Beiträge aus Forschung und Praxis* (S. 17-49). Neuwied: Luchterhand.

Lewin, K., Heublein, U., Schreiber, J., Spangenberg, H. & Sommer, D. (2002). *Studienanfänger im Wintersemester 2000/2001: Trotz Anfangsschwierigkeiten optimistisch in die Zukunft*. Hannover: Hochschul-Informations-System.

- Lewin, K., Heublein, U. & Sommer, D. (2000). *Differenzierung und Homogenität beim Hochschulzugang (A7/2000)*. Hannover: Hochschul-Informations-System.
- Licht, G. (2002). *IKT-Fachkräftemangel und Qualifikationsbedarf*. Baden-Baden: Nomos.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Psychologie-Verlags-Union.
- Lissmann, U. (1977). Vorhersagewert gewichteter Reifezeugnisse. *Zeitschrift für Pädagogische Psychologie*, 23, 107-118.
- Long, L. & Tracey, T. J. G. (2006). Structure of RIASEC scores in China: A structural meta-analysis. *Journal of Vocational Behavior*, 68, 39-51.
- Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "General Intelligence,' objectively determined and measured". *Journal of Personality and Social Psychology*, 86, 96-111.
- Lubinski, D. & Humphreys, L. G. (1997). Incorporating general intelligence into epidemiology and the social sciences. *Intelligence*, 24, 159-201.
- Lückefett, H. J. & Thomas, U. (2001). *Die Entwicklung des Arbeitsmarktes und der Hochschulplätze für IT-Fachkräfte in Deutschland*. Frankfurt am Main: Initiative D21.
- Lutz, B. & Meil, P. (2000). Thesen zum zukünftigen Qualifikationsbedarf der deutschen Industrie. In B. Lutz, P. Meil & B. Wiener (Hrsg.), *Industrielle Fachkräfte für das 21. Jahrhundert. Aufgaben und Perspektiven für die Produktion von morgen* (S. 19-38). Frankfurt am Main: Campus.
- Mackintosh, N. J. & Bennett, E. S. (2005). What do Raven's matrices measure? An analysis in terms of sex differences. *Intelligence*, 33, 663-674.
- Marshalek, B., Lohman, D. F. & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence*, 7, 107-127.
- Maurice, J. v., Scheller, R. & Bäumer, T. (1995). Zufallserfahrungen und Interessenstruktur: Eine Untersuchung zum Wahlverhalten von Studienanfängern. *Zeitschrift für Pädagogische Psychologie*, 9, 37-44.
- McNulty, J. L. (2000). Five-factor model of personality. In A. E. Kazdin (Hrsg.), *Encyclopedia of psychology*, Vol. 3 (S. 375-376). Washington, DC: American Psychological Association.
- Mead, A. D. & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114, 449-458.

-
- Meier, B. (2003). Ist der Erfolg im Jurastudium vorhersagbar? Empirische Befunde zum Zusammenhang zwischen Schulnoten und Abschneiden im Ersten Juristischen Staatsexamen. *Beiträge zur Hochschulforschung*, 25, 18-36.
- Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., Eisman, E. J., Kubiszyn, T. W. & Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, 56, 128-165.
- Middleton, J. W., Tate, R. L. & Geraghty, T. J. (2003). Self-Efficacy and spinal cord injury: Psychometric properties of a new scale. *Rehabilitation Psychology*, 48, 281-288.
- Moosbrugger, H., Jonkisz, E. & Fucks, S. (2006). Studierendenauswahl durch die Hochschulen - Ansätze zur Prognostizierbarkeit des Studienerfolgs am Beispiel des Studiengangs Psychologie. *Report Psychologie*, 31, 114-123.
- Morrison, D. G. (1969). On the interpretation of discriminant analysis. *Journal of Marketing Research*, 6, 156-163.
- Mullis, I. V., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L. & Smith, T. A. (1998). *Mathematics and science achievement in the final year of secondary school*. Chestnut Hill, MA: TIMSS International Study Center.
- Multon, K. D., Brown, S. D. & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analysis investigation. *Journal of Counseling Psychology*, 38, 30-38.
- Naglieri, J. A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A. & Velasquez, R. (2004). Psychological testing on the internet - New problems, old issues. *American Psychologist*, 59, 150-162.
- Nerdinger, F. W. (2001). Motivierung. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 349-372). Göttingen: Hogrefe.
- Newstead, S. E. (1992). A study of two 'quick-and-easy' methods of assessing individual differences in student learning. *British Journal of Educational Psychology*, 62, 299-312.
- Nielsen-NetRatings (2003). *Global internet population grows an average of four percent year-over-year*. Online in Internet: URL: http://www.nielsen-netratings.com/pr/pr_030220.pdf (Stand 3. Januar 2005).
- Nielsen-NetRatings (2004). *Europe is at home online: 100 million and counting*. Online in Internet: URL: http://direct.www.nielsen-netratings.com/pr/pr_041202_uk.pdf (Stand 25. Dezember 2004).
- Noel-Levitz (2005). *Retention Revenue Estimator*. Online in Internet: URL: <http://www.noellevitz.com/Papers+and+Research/Retention+Calculator/> (Stand 7. März 2006).

- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments und Computers*, 32, 396-402.
- Offer, M. & Sampson, J. P. (1999). Quality in the content and use of information and communications technology in guidance. *British Journal of Guidance and Counselling*, 27, 501-516.
- Olea, M. M. & Ree, M. J. (1994). Predicting pilot and navigator criteria: Not much more than g. *Journal of Applied Psychology*, 79, 845-851.
- Oliver, L. W. & Spokane, A. R. (1988). Career-intervention outcome: What contributes to client gain? *Journal of Counseling Psychology*, 35, 447-462.
- Oliver, L. W. & Zack, J. S. (1999). Career assessment on the internet: An exploratory study. *Journal of Career Assessment*, 7, 323-356.
- Organisation for Economic Co-operation and Development (2002). *OECD review of career guidance policies - Germany*. Online in Internet: URL: <http://www.oecd.org/dataoecd/28/35/1939079.pdf> (Stand 3. Februar 2005).
- Organisation für Wirtschaftliche Zusammenarbeit und Entwicklung (2004). *Bildung auf einen Blick - OECD-Indikatoren 2004*. Bielefeld: Bertelsmann.
- Organisation für Wirtschaftliche Zusammenarbeit und Entwicklung (2005). *Bildung auf einen Blick - OECD-Indikatoren 2005*. Bielefeld: Bertelsmann.
- Organisation für Wirtschaftliche Zusammenarbeit und Entwicklung (2006). *Bildung auf einen Blick - OECD-Indikatoren 2006*. Bielefeld: Bertelsmann.
- Osburn, H. (1996). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods*, 5, 343-355.
- Parsons, F. (1909). *Choosing a vocation*. Boston: Houghton Mifflin.
- Pascarella, E. T. & Chapman, D. W. (1983). A multiinstitutional, path analytic validation of Tinto's model of college withdrawal. *American Educational Research Journal*, 20, 87-102.
- Paterson, D. G., Schneidler, G. G. & Williamson, E. G. (1938). Scholastic aptitude tests. In D. G. Paterson, G. G. Schneidler & E. G. Williamson (Hrsg.), *Student guidance techniques: A handbook for counselors in high schools and colleges* (S. 52-94). New York, NY: McGraw-Hill Book Company.
- Pawlik, K. (1971). Concepts in human cognition and aptitudes. In R. B. Cattell (Hrsg.), *Handbook of multivariate experimental psychology* (S. 535-562). Boston, MA: Houghton Mifflin.

-
- Pekrun, R. (1988). *Emotion, Motivation und Persönlichkeit*. Weinheim: Psychologie-Verlags-Union.
- Petersen, A. & Wehmeyer, C. (2001). *Die neuen IT-Berufe auf dem Prüfstand. Bestand sowie Bedarf und Angebot an IT-Fachkräften*. Flensburg: Berufsbildungsinstitut Arbeit und Technik Universität Flensburg.
- Peterson, N. G., Mumford, M. D., Borman, W. C., Jeanneret, P. R., Fleishman, E. A., Levin, K. Y., Campion, M. A., Mayfield, M. S., Morgeson, F. P., Pearlman, K., Gowing, M. K., Lancaster, A. R. & Silve, (2001). Understanding work using the occupational information network (O*NET). *Personnel Psychology*, 54, 451-492.
- Picht, G. (1965). *Die deutsche Bildungskatastrophe*. München: Deutscher Taschenbuch-Verlag.
- Pitz, G. F. & Harren, V. A. (1980). An analysis of career decision making from the point of view of information processing and decision theory. *Journal of Vocational Behavior*, 16, 320-346.
- Prediger, D. J. (1982). Dimensions underlying Holland's hexagon: Missing link between interests and occupations. *Journal of Vocational Behavior*, 21, 259-287.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rolff, H., Rost, J. & Schiefele, U. (2004). *PISA 2003 – Ergebnisse des zweiten internationalen Vergleichs - Zusammenfassung*. Online in Internet: URL: http://pisa.ipn.uni-kiel.de/Ergebnisse_PISA_2003.pdf (Stand 8. Mai 2005).
- Prenzel, M., Krapp, A. & Schiefele, H. (1986). Grundzüge einer pädagogischen Interessentheorie. *Zeitschrift für Pädagogik*, 32, 163-173.
- Rahn, H. (1976). *Testentwicklung und Testeinsatz durch die Studienstiftung des deutschen Volkes*. Bonn: Institut für Test- und Begabungsforschung.
- Rammstedt, B., Koch, K., Borg, I. & Reitz, T. (2004). Entwicklung und Validierung einer Kurzskala für die Messung der Big-Five-Persönlichkeitsdimensionen in Umfragen. *Zusammenhänge*, 55, 5-28.
- Raven, J. C. (1989). The Raven progressive matrices: A review of national norming studies and ethnic and socioeconomic variation within the United States. *Journal of Educational Measurement*, 26, 1-16.
- Raven, J. C. (2004). *Advanced progressive matrices*. Oxford: Oxford Psychologists Press.
- Ree, M. J. & Carretta, T. R. (2002). g2K. *Human Performance*, 15, 3-24.

- Ree, M. J. & Earles, J. A. (1991). Predicting training success: Not much more than g. *Personnel Psychology*, 44, 321-332.
- Ree, M. J., Earles, J. A. & Teachout, M. S. (1994). Predicting job performance: Not much more than g. *Journal of Applied Psychology*, 79, 518-524.
- Reinberg, A. & Hummel, M. (2003a). *In der Krise verdrängt, sogar im Boom vergessen*. Online in Internet: URL: <http://doku.iab.de/kurzber/2003/kb1903.pdf> (Stand 7. April 2005).
- Reinberg, A. & Hummel, M. (2003b). Steuert Deutschland langfristig auf einen Fachkräftemangel zu? *IAB Kurzberichte*, 9, 1-8.
- Reinberg, A. & Hummel, M. (2005). Vertrauter Befund: Höhere Bildung schützt auch in der Krise vor Arbeitslosigkeit. *Institut für Arbeitsmarkt- und Berufsforschung*, 9, 1-6.
- Reinberg, A. & Schreyer, F. (2003). *Arbeitsmarkt für AkademikerInnen - Studieren lohnt sich auch in der Zukunft*. Online in Internet: URL: <http://doku.iab.de/kurzber/2003/kb2003.pdf> (Stand 7. April 2005).
- Reips, U. (2002). Standards for internet-based experimenting. *Experimental Psychology*, 49, 243-256.
- Reissert, R. (1983). *Studienabbruch im Widerstreit von Ergebnissen und Meinungen (HIS-Kurzinformationen A1)*. Hannover: Hochschul-Informations-System.
- Reissert, R. (1991). Fachstudiendauer: Ist das Problem schon fixiert, welche Handlungsmöglichkeiten gibt es? In W. D. Webler & H. U. Otto (Hrsg.), *Der Ort der Lehre in der Hochschule* (S. 29-60). Weinheim: Deutscher Studien Verlag.
- Renner, K., Marcus, B., Machilek, F. & Schütz, A. (2005). Selbstdarstellung und Persönlichkeit auf privaten Homepages. In K. Renner, A. Schütz & F. Machilek (Hrsg.), *Internet und Persönlichkeit* (S. 189-204). Göttingen: Hogrefe.
- Rettig, K., Hornke, L. F. & Schiff, B. (1989). Konstruktion eines Rechentests. *Untersuchungen des Psychologischen Dienstes der Bundeswehr*, 24, 139-201.
- Rheinberg, F. (1995). *Motivation*. Stuttgart: Kohlhammer.
- Rheinberg, F. (2006). Intrinsische Motivation und Flow-Erleben. In J. Heckhausen & H. Heckhausen (Hrsg.), *Motivation und Handeln* (S. 331-354). Berlin: Springer.
- Richman, W. L., Kiesler, S., Weisband, S. & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology*, 84, 754-775.

-
- Rindermann, H. (2005). Für ein bundesweites Auswahlverfahren von Studienanfängern über Fähigkeitsmessung. *Psychologische Rundschau*, 56, 127-129.
- Rindermann, H. & Oubaid, V. (1999). Auswahl von Studienanfängern durch Universitäten - Kriterien, Verfahren und Prognostizierbarkeit des Studienerfolgs. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 20, 172-191.
- Rips, L. J. & Conrad, F. G. (1983). Individual differences in deduction. *Cognition and Brain Theory*, 6, 259-285.
- Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R. & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, 130, 261-288.
- Roberts, J. K. & Henson, R. K. (2002). Correction for bias in estimating effect sizes. *Educational and Psychological Measurement*, 62, 241-253.
- Rohracher, H. (1988). *Einführung in die Psychologie* (13. Aufl.). Weinheim: Psychologie-Verlags-Union.
- Rolfs, H. (2001). *Berufliche Interessen*. Göttingen: Hogrefe.
- Rolfs, H. & Schuler, H. (2002a). Berufliche Interessenkongruenz und das Erleben im Studium. *Zeitschrift für Arbeits- und Organisationspsychologie*, 46, 137-149.
- Rolfs, H. & Schuler, H. (2002b). Lernstrategien im Studium als Korrelat beruflicher Interessenkongruenz. *Psychologie in Erziehung und Unterricht*, 49, 250-262.
- Rosas Díaz, R. (1990). *Replikation des Berliner Intelligenzstrukturmodells (BIS) und Vorhersagbarkeit des Studienerfolgs bei chilenischen Studenten*. Unveröffentlichte Dissertation, Freie Universität Berlin.
- Rosenfeld, P., Booth-Kewley, S., Edwards, J. & Thomas, M. (1996). Responses on computer surveys: Impression management, social desirability, and the Big Brother syndrome. *Computers in Human Behavior*, 12, 263-274.
- Rounds, J., McKenna, M. C., Hubert, L. & Day, S. X. (2000). Response Tinsley on Holland: A misshapen argument. *Journal of Vocational Behavior*, 56, 205-215.
- Rounds, J. & Tracey, T. J. (1996). Cross-cultural structural equivalence of RIASEC models and measures. *Journal of Counseling Psychology*, 43, 310-329.
- Rounds, J., Tracey, T. J. & Hubert, L. (1992). Methods for evaluating vocational interest structural hypotheses. *Journal of Vocational Behavior*, 40, 239-259.

- Ryan, J. M., Tracey, T. J. & Rounds, J. (1996). Generalizability of Holland's structure of vocational interests across ethnicity, gender, and socioeconomic status. *Journal of Counseling Psychology*, 43, 330-337.
- Sadler-Smith, E. (1997). 'Learning style': Frameworks and instruments. *Educational Psychology*, 17, 51-63.
- Salgado, J. F. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology*, 82, 30-43.
- Salgado, J. F. (1998). Big Five personality dimensions and job performance in Army and civil occupations: A European perspective. *Human Performance*, 11, 271-288.
- Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C. & de Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities: A european community meta-analysis. *Personnel Psychology*, 56, 573-605.
- Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., de Fruyt, F. & Rolland, J. P. (2003). A meta-analytic study of general mental ability validity for different occupations in the european community. *Journal of Applied Psychology*, 88, 1068-1081.
- Salgado, J. F. & Moscoso, S. (2003). Internet-based personality testing: Equivalence of measure and assess' perceptions and reactions. *International Journal of Selection and Assessment*, 11, 194-205.
- Sampson, J. P. J., Purgar, M. P. & Shy, J. D. (2003). Computer-based test interpretation in career assessment: Ethical and professional issues. *Journal of Career Assessment*, 11, 22-39.
- Schiefele, U., Krapp, A. & Schreyer, I. (1993). Metaanalyse des Zusammenhangs von Interesse und schulischer Leistung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, XXV, 120-148.
- Schiefele, U. & Moschner, B. & H. R. (2002). *Skalenhandbuch SMILE-Projekt*. Universität Bielefeld, Abteilung für Psychologie.
- Schiefele, U. & Schreyer, I. (1994). Intrinsische Lernmotivation und Lernen. Ein Überblick zu Ergebnissen der Forschung. *Zeitschrift für Pädagogische Psychologie*, 8, 1-13.
- Schiefele, U. & Urhahne, D. (2000). Motivationale und volitionale Bedingungen der Studienleistung. In U. Schiefele & K. Wild (Hrsg.), *Interesse und Lernmotivation* (S. 183-206). Münster: Waxmann.
- Schiefele, U., Wild, K. & Winteler, A. (1995). Lernaufwand und Elaborationsstrategien als Mediatoren der Beziehung von Studieninteresse und Studienleistung. *Zeitschrift für Pädagogische Psychologie*, 9, 181-188.

-
- Schiefele, U., Winteler, A. & Krapp, A. (1988). Studieninteresse und fachbezogene Wissenstrukturen. *Psychologie in Erziehung und Unterricht*, 35, 106-118.
- Schindler, G. (1997). *"Frühe" und "späte" Studienabbrecher*. München: Bayerisches Staatsinstitut für Hochschulforschung und Hochschulplanung.
- Schindler, G. (1999). Fallstudien zum Studienabbruch: Frühe und späte Studienabbrecher. In M. Schröder-Gronostay & H. Daniel (Hrsg.), *Studienerfolg und Studienabbruch* (S. 161-179). Neuwied: Luchterhand.
- Schmidt, F. L. (1992). What do you really mean? Researches, Meta-Analysis, and Cumulative Knowledge in Psychology. *American Psychologist*, 47, 1173-1181.
- Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance*, 15, 187-211.
- Schmidt, F. L. & Hunter, J. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.
- Schmidt, F. L. & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, 86, 162-173.
- Schmidt-Atzert, L., Deter, B., Jaeckel Deter; Jaeckel, & Jaeckel, S. (2004). Prädiktion von Ausbildungserfolg: Allgemeine Intelligenz (g) oder spezifische kognitive Fähigkeiten? *Zeitschrift für Personalpsychologie*, 3, 147-158.
- Schmidt-Klingenberg, M. (2002). *Was ist das Abitur noch wert?* Online in Internet: URL: <http://www.spiegel.de/unispiegel/schule/0,1518,202526,00.html> (Stand 6. August 2006).
- Schmitt, C. & Winkelmann, U. (2005). *Wer bleibt kinderlos?* Berlin: German institute for economic research.
- Schmitt, N., Oswald, F. L., Kim, B. H., Imus, A., Merritt, S., Friede, A. & Shivpuri, S. (2007). The use of background and ability profiles to predict college student outcomes. *Journal of Applied Psychology*, 92, 165 - 179.
- Scholz, U., Dona, B. G., Sud, S. & Schwarzer, R. (2002). Is general self-efficacy a universal construct? Psychometric findings from 25 countries. *European Journal of Psychological Assessment*, 18, 242-251.
- Schröder-Gronostay, M. (1999). Studienabbruch - Zusammenfassung des Forschungsstandes. In M. Schroeder-Gronostay & H. Daniel (Hrsg.), *Studienerfolg und Studienabbruch* (S. 209-240). Neuwied: Luchterhand.

- Schröder-Gronostay, M. & Daniel, H. (1999). Vorwort. In M. Schröder-Gronostay & H. Daniel (Hrsg.), *Studienerfolg und Studienabbruch: Beiträge aus Forschung und Praxis* (S. VII-XI). Neuwied: Luchterhand.
- Schuler, H. (1990). Personalauswahl aus der Sicht der Bewerber: Zum Erleben eignungsdiagnostischer Situationen. *Zeitschrift für Arbeit- und Organisationspsychologie*, 34, 184-191.
- Schuler, H. (1996). *Psychologische Personalauswahl*. Göttingen: Verlag für Angewandte Psychologie.
- Schuler, H. (1998). Noten und Studien- und Berufserfolg. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 370-374). Weinheim: Psychologie-Verlags-Union.
- Schuler, H., Funke, U. & Baron-Boldt, J. (1990). Predictive validity of school grades: A meta-analysis. *Applied Psychology: An International Review*, 39, 89-103.
- Schuler, H. & Prochaska, M. (2001). *Leistungsmotivationsinventar: LMI*. Göttingen: Hogrefe.
- Sedlacek, G. (2004). *Analyse der Studiendauer und des Studienabbruch-Risikos: unter Verwendung der statistischen Methoden der Ereignisanalyse*. Frankfurt am Main: Lang.
- Sedlacek, W. E. (2001). *Why we should use noncognitive variables with graduate students*. Online in Internet: URL: <http://eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED456717> (Stand 25. März 2006).
- Sedlacek, W. E. (2004). *Beyond the big test: Noncognitive assessment in higher education*. San Francisco, CA: Jossey-Bass.
- Seidman, A. (2005). *College student retention*. Westport, CT: Praeger.
- Shea, D. L., Lubinski, D. & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, 93, 604-614.
- Silvia, P. J. (2001). Expressed and measured vocational interests: Distinctions and definitions. *Journal of Vocational Behavior*, 59, 382-393.
- Simeaner, H., Röhl, T. & Bargel, T. (2004). *Studiensituation und Studierende - Datenalmanach - Studierendensurvey 1983 - 2001*. Konstanz: Arbeitsgruppe Hochschulforschung, Universität Konstanz.
- Spady, W. G. (1970). Dropouts from higher education: An interdisciplinary review and synthesis. *Interchange*, 1, 64-85.
- Spearman, C. (1904). "General Intelligence," objectively determined and measured. *American Journal of Psychology*, 15, 201-293.

-
- Spearman, C. (1927). *The abilities of man*. New York City, NY: Macmillan Company.
- Spiegel-Online (2001). *Das Land der Dichter und Denker - abgehängt*. Online in Internet: URL: <http://www.spiegel.de/unispiegel/studium/0,1518,172574,00.html> (Stand 8. April 2005).
- Spiegel-Online (2002). *Trotz Abi nicht fit für die Uni?* Online in Internet: URL: <http://www.spiegel.de/unispiegel/studium/0,1518,176556,00.html> (Stand 10. Oktober 2005).
- Spies, C. (1999). Der Studienfachwechsel - Vorbote für einen Abbruch oder Ausgangspunkt für einen erfolgreichen Abschluss? In M. Schröder-Gronostay & H. Daniel (Hrsg.), *Studienerfolg und Studienabbruch* (S. 181-193). Neuwied: Luchterhand.
- Spies, K., Westermann, R., Heise, E. & Hagen, M. (1998). Zur Abhängigkeit von Studienzufriedenheit von Diskrepanzen zwischen Fähigkeiten und Anforderungen. *Psychologie in Erziehung und Unterricht*, 45, 36-52.
- Spiewak, M. (2004). Wettbewerb lebt von Unterschieden. *Aus Politik und Zeitgeschichte*, 25, 3-5.
- Spokane, A. R. (1985). A review of research on person-environment congruence in Holland's theory of careers. *Journal of Vocational Behavior*, 26, 306-343.
- Spokane, A. R., Meir, E. I. & Catalano, M. (2000). Person-Environment congruence and Holland's Theory: A review and reconsideration. *Journal of Vocational Behavior*, 57, 137-187.
- SPSS Inc. (2003). SPSS 12.0 für Windows [Computersoftware]. Chicago, IL: Autor.
- Spychalski, A. C., Quinones, M. A., Gaugler, B. B. & Pohley, K. (1997). A survey of assessment center practices in organizations in the United States. *Personnel Psychology*, 50, 71-90.
- Statistisches Bundesamt (2001). *Hochschulstandort Deutschland 2001*. Wiesbaden: Autor.
- Statistisches Bundesamt (2005a). *Bildung und Kultur - Nichtmonetäre hochschulstatistische Kennzahlen - 1980-2003*. Wiesbaden: Autor.
- Statistisches Bundesamt (2005b). *Prüfungen an Hochschulen*. Wiesbaden: Autor.
- Statistisches Bundesamt (2006). *Bildung und Kultur - Schnellmeldungsergebnisse der Hochschulstatistik zu Studierenden und Studienanfänger/-innen - vorläufige Ergebnisse*. Wiesbaden: Autor.
- Stebler, P. (2000). *Studienerfolg und Studienzufriedenheit an der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität Freiburg i.Ue.* Unveröffentlichte Dissertation, Universität Freiburg/Schweiz.

- Stegmann, H. (1996). Zum Zusammenhang zwischen Schulnoten und Berufserfolg. In B. Ertelt & M. Hofer (Hrsg.), *Theorie und Praxis der Beratung* (S. 161-183). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung der Bundesanstalt für Arbeit.
- Stemmler, G. (2005). Studierendenauswahl durch Hochschulen: Ungewisser Nutzen. *Psychologische Rundschau*, 56, 125-127.
- Steptoe, A., Perkins-Porras, L., Rink, E., Hilton, S. & Cappuccio, F. P. (2004). Psychological and social predictors of changes in fruit and vegetable consumption over 12 months following behavioral and nutrition education counseling. *Health Psychology*, 23, 574-581.
- Sternberg, R. J. & Williams, W. M. (1997). Does the Graduate Record Examination predict meaningful success in the graduate training of psychologists? A case study. *American Psychologist*, 52, 630-641.
- Stichting LogReport Foundation (2004). Lire [Computersoftware]. Eindhoven: Autor.
- Stoecker, J., Pascarella, E. T. & Wolfe, L. M. (1988). Persistence in higher education: A 9-year test of a theoretical model. *Journal of College Student Development*, 29, 196-209.
- Stumpf, H. & Nauels, H. (1990). Zur prognostischen Validität des "Tests für medizinische Studiengänge" (TMS) im Studiengang Humanmedizin. *Diagnostica*, 36, 16-32.
- Süß, H. (2001). Prädiktive Validität der Intelligenz im schulischen und außerschulischen Bereich. In E. Stern & J. Guthke (Hrsg.), *Perspektiven der Intelligenzforschung* (S. 109-135). Lengerich: Pabst Science Publisher.
- Süß, H. (2003). Intelligenztheorien. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 217-224). Weinheim: Psychologie-Verlags-Union.
- Taylor, H. C. & Russell, J. T. (1939). The relationship of validity coefficients to the practical validity of tests in selection: Discussion and tables. *Journal of Applied Psychology*, 23, 565-578.
- Teichler, U. (2005). Quantitative und strukturelle Entwicklungen des Hochschulwesens. *Zeitschrift für Pädagogik*, 51, 8-24.
- Tett, R. P., Jackson, D. N. & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, 44, 703-742.
- Tinsley, H. E. A. (2000a). The congruence myth: An analysis of the efficacy of person-environment fit model. *Journal of Vocational Behavior*, 56, 147-179.
- Tinsley, H. E. A. (2000b). The congruence myth revisited. *Journal of Vocational Behavior*, 56, 405-423.

-
- Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (2. Aufl.). Chicago, IL: University of Chicago Press.
- Todt, E. (1978). *Das Interesse: Empirische Untersuchungen zu einem Motivationskonzept*. Bern: Huber.
- Tracey, T. J. G. & Rounds, J. (1993). Evaluating Holland's and Gati's vocational-interest models: A structural meta-analysis. *Psychological Bulletin*, 113, 229-246.
- Tracey, T. J. G. & Hopkins, N. (2001). Correspondence of interests and abilities with occupational choice. *Journal of Counseling Psychology*, 48, 178-189.
- Tracey, T. J. G., Darcy, M. & Kovalski, T. M. (2000). A closer look at person-environment fit. *Journal of Vocational Behavior*, 56, 216-224.
- Tracey, T. J. G., Watanabe, N. & Schneider, P. L. (1997). Structural invariance of vocational interests across Japanese and American cultures. *Journal of Counseling Psychology*, 44, 346-354.
- Tranberg, M., Slane, S. & Ekeberg, S. E. (1993). The relation between interest congruence and satisfaction: A metaanalysis. *Journal of Vocational Behavior*, 42, 253-264.
- Trapmann, S., Hell, B., Hirn, J. O., Weigand, S. & Schuler, H. (2005, September). *Psychologische Konstrukte als Prädiktoren des Studienerfolgs - eine Metaanalyse*. Vortrag auf der 8. Arbeitstagung Differentielle Psychologie, Persönlichkeitspsychologie und Psychologische Diagnostik, Marburg.
- Trapmann, S., Hell, B., Weigand, S. & Schuler, H. (2007). Die Validität von Schulnoten zur Vorhersage des Studienerfolgs - eine Metaanalyse. *Zeitschrift für Pädagogische Psychologie*, 21, 11-27.
- Trost, G. (1975). *Vorhersage des Studienerfolgs*. Braunschweig: Westermann.
- Trost, G. (1996). Interview. In K. Pawlik (Hrsg.), *Grundlagen und Methoden der Differentiellen Psychologie* (S. 464-505). Göttingen: Hogrefe.
- Trost, G. (2003). *Deutsche und internationale Studierfähigkeitstests: Arten, Brauchbarkeit, Handhabung*. Bonn: Deutscher Akademischer Austauschdienst.
- Trost, G. & Haase, K. (2005). *Hochschulzulassung: Auswahlmodelle für die Zukunft*. Stuttgart: Landesstiftung Baden-Württemberg.
- Trost, G., Klieme, E. & Nauels, H. (1997). Prognostische Validität des Tests für medizinische Studiengänge (TMS). In T. Herrmann (Hrsg.), *Hochschulentwicklung - Aufgaben und Chancen* (S. 57-78). Heidelberg: Asanger.

- Trouteaud, A. R. (2004). How you ask counts: A test of internet-related components of response rates to a web-based survey. *Social Science Computer Review*, 22, 385-392.
- Tsabari, O., Tziner, A. & Meir, E. I. (2005). Updated meta-analysis on the relationship between congruence and satisfaction. *Journal of Career Assessment*, 13, 216-232.
- Urhahne, D. (1997). *Interesse, Volition und Lernleistung - Motivationale Faktoren erfolgreichen Lernens*. Unveröffentlichte Diplomarbeit, Universität Bielefeld.
- Vacha-Haase, T. & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology*, 51, 473-481.
- Vernon, P. A. (1987). *Speed of information-processing and intelligence*. Norwood, NJ: Ablex.
- Viebahn, P. (1990). *Psychologie des studentischen Lernens*. Weinheim: Deutscher Studien Verlag.
- Volet, S. E. (1997). Cognitive and affective variables in academic learning: The significance of direction and effort in students' goals. *Learning and Instruction*, 7, 235-254.
- Waterman, A. S. & Waterman, C. K. (1972). Relationship between freshman ego identity status and subsequent academic behavior: A test of the predictive validity of Marcia's categorization system of identity status. *Developmental Psychology*, 6, 179.
- Weber, V. (2003). *Entwicklung und erste Erprobung eines web-basierten Self Assessments zur Feststellung der Studieneignung für die Fächer Elektrotechnik, Technische Informatik sowie Informatik an der RWTH Aachen*. Unveröffentlichte Diplomarbeit, RWTH Aachen.
- Weck, M. (1991). *Der Studienfachwechsel*. Frankfurt am Main: Peter Lang.
- Weinert, A. B. (1992). *Lehrbuch der Organisationspsychologie* (3. Aufl.). Weinheim: Psychologie-Verlags-Union.
- Westermann, R., Heise, E., Spies, K. & Trautwein, U. (1996). Identifikation und Erfassung von Komponenten der Studienzufriedenheit. *Psychologie in Erziehung und Unterricht*, 43, 1-22.
- Whiston, S. C., Sexton, T. L. & Lasoff, D. L. (1998). Career-intervention outcome: A replication and extension of Oliver and Spokane (1988). *Journal of Counseling Psychology*, 45, 150-165.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48, 817-838.
- Wiggins, J. S. (1973). *Personality and prediction*. Reading, MA: Addison-Wesley.
- Wild, B., De la Fontaine, A. & Schafsteller, C. (2001). Fishing for Talents: Internet-Recruiting auf neuen Wegen. *Personalführung*, 1, 66-70.

-
- Wild, K. (2001). Lernstrategien und Lernstile. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 424-429). Weinheim: Psychologie-Verlags-Union.
- Wild, K., Krapp, A. & Winteler, A. (1992). Die Bedeutung von Lernstrategien zur Erklärung des Einflusses von Studieninteresse auf Lernleistungen. In A. Krapp & M. Prenzel (Hrsg.), *Interesse, Lernen, Leistung: Neuere Ansätze der pädagogisch-psychologischen Interessenforschung* (S. 279-295). Münster: Aschendorff.
- Wilk, S. L., Desmarais, L. B. & Sackett, P. R. (1995). Gravitation to jobs commensurate with ability: Longitudinal and cross-sectional tests. *Journal of Applied Psychology*, 80, 79-85.
- Williams, R. H., Zimmerman, D. W., Zumbo, B. D. & Ross, D. (2004). Charles Spearman: British behavioral scientist. *Human Nature Review*, 3, 114-118.
- Winteler, A., Sierwald, W. & Schiefele, U. (1988). Interesse, Leistung und Wissen: Die Erfassung von Studieninteresse und seine Bedeutung für Studienleistung und fachbezogenes Wissen. *Empirische Pädagogik*, 2, 227-250.
- Wissenschaftsrat (1999). *Stellungnahme zum Verhältnis von Hochschulausbildung und Beschäftigungssystem*. Würzburg: Autor.
- Wissenschaftsrat (2001). *Entwicklung der Fachstudiendauer an Hochschulen von 1990 bis 1998*. Online in Internet: URL: <http://www.wissenschaftsrat.de/texte/4770-01.pdf> (Stand 18. April 2004).
- Wissenschaftsrat (2003). *Prüfungsnoten an Hochschulen 1996, 1998 und 2000 nach ausgewählten Studienbereichen und Studienfächern*. Online in Internet: URL: <http://www.wissenschaftsrat.de/texte/5526-03-pdf> (Stand 18. April 2004).
- Wissenschaftsrat (2004). *Empfehlungen zur Reform des Hochschulzugangs*. Online in Internet: URL: <http://www.wissenschaftsrat.de/texte/5920-04.pdf> (Stand 18. April 2004).
- Wittmann, W. W. (2002). Brunswik-Symmetrie: Ein Schlüsselkonzept für erfolgreiche psychologische Forschung. In M. Myrtek (Hrsg.), *Die Person im biologischen und sozialen Kontext* (S. 163-186). Göttingen: Hogrefe.
- Wittmann, W. W. & Matt, G. E. (1986). Aggregation und Symmetrie. Grundlagen einer multivariaten Reliabilitäts- und Validitätstheorie, dargestellt am Beispiel der differentiellen Validität des Berliner Intelligenzstrukturmodells. *Diagnostica*, 32, 309-329.
- Wittmann, W. W. & Süß, H. (1999). Learning and individual differences: Process, trait, and content determinants. In P. L. Ackerman, P. C. Kyllonen & R. D. Roberts (Hrsg.), *Learning and individual differences: Process, trait, and content determinants* (S. 77-108). Washington, DC: American Psychological Association.

- Wolfe, R. N. & Johnson, S. D. (1995). Personality as a predictor of college performance. *Educational and Psychological Measurement*, 55, 177-185.
- Wolter, A. (1989). *Von der Elitenbildung zur Bildungsexpansion*. Oldenburg: Bibliotheks- und Informationssystem der Universität Oldenburg.
- Yeo, G. B. & Neal, A. (2004). A multilevel analysis of effort, practice, and performance: Effects of ability, conscientiousness, and goal orientation. *Journal of Applied Psychology*, 89, 231-247.
- Zimmerhofer, A. (2003). *Neukonstruktion und erste Erprobung eines webbasierten SelfAssessments zur Feststellung der Studienneigung für die Fächer Elektrotechnik, Technische Informatik sowie Informatik an der RWTH Aachen*. Unveröffentlichte Diplomarbeit, RWTH Aachen.
- Zimmerhofer, A., Heukamp, V. & Hornke, L. F. (2006). Ein Schritt zur fundierteren Studienfachwahl - webbasierte Self-Assessments in der Praxis. *Report Psychologie*, 31, 62-72.
- Zimmerhofer, A. & Hornke, L. F. (2005a, September). *Data-Screening von großen Datensätzen bei der Interneterhebung zur Studienberatung*. Vortrag auf der 8. Arbeitstagung der Fachgruppe für Differentielle Psychologie, Persönlichkeitspsychologie und Psychologische Diagnostik, Marburg.
- Zimmerhofer, A. & Hornke, L. F. (2005b). Konzeption einer webbasierten Studienberatung für Interessierte der Studienfächer Informatik, Elektrotechnik und Technische Informatik. In K. Renner, A. Schütz & F. Machilek (Hrsg.), *Internet und Persönlichkeit* (S. 269-284). Göttingen: Hogrefe.
- Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25, 82-91.
- Zweites Deutsches Fernsehen (2005). *Lustlose Lehrlinge - Das Elend mit den Auszubildenden*. Online in Internet: URL: <http://www.zdf.de/ZDFde/inhalt/8/0,1872,2347464,00.html> (Stand 24. März 2006).
- Zwick, M. M. & Renn, O. (2000). *Die Attraktivität von technischen und ingenieurwissenschaftlichen Fächern bei der Studien- und Berufswahl junger Frauen und Männer*. Stuttgart: Akademie für Technikfolgenabschätzung in Baden-Württemberg.

Kapitel 11

Anhang

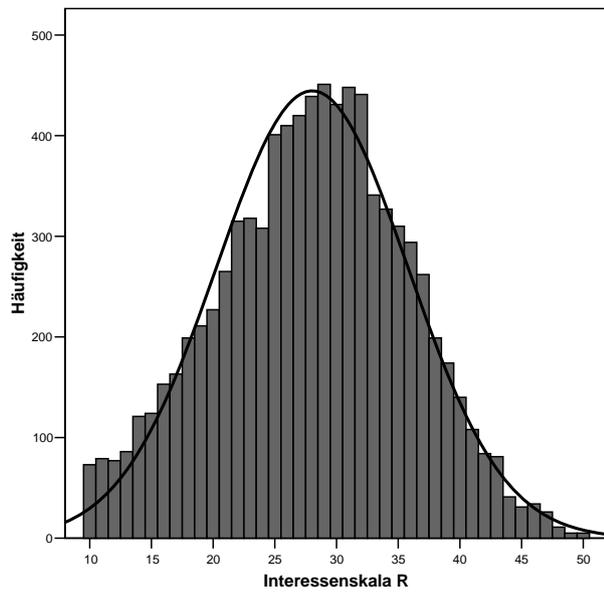


Abbildung 11.1: Scoreverteilung Skala INT-R

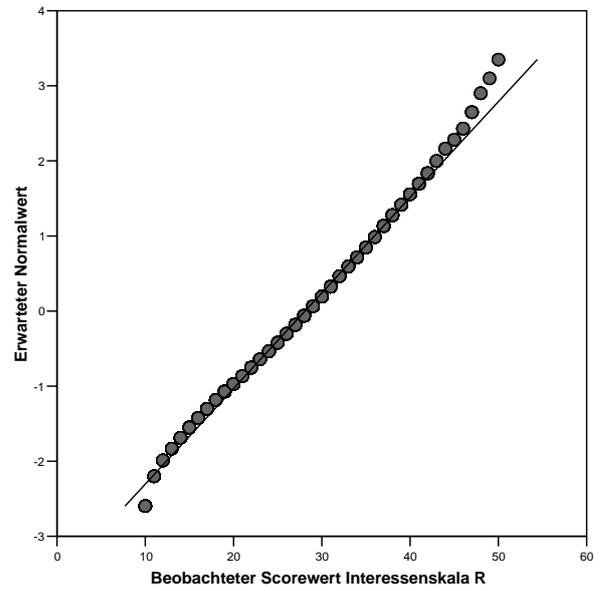


Abbildung 11.2: Q-Q-Diagramm Skala INT-R

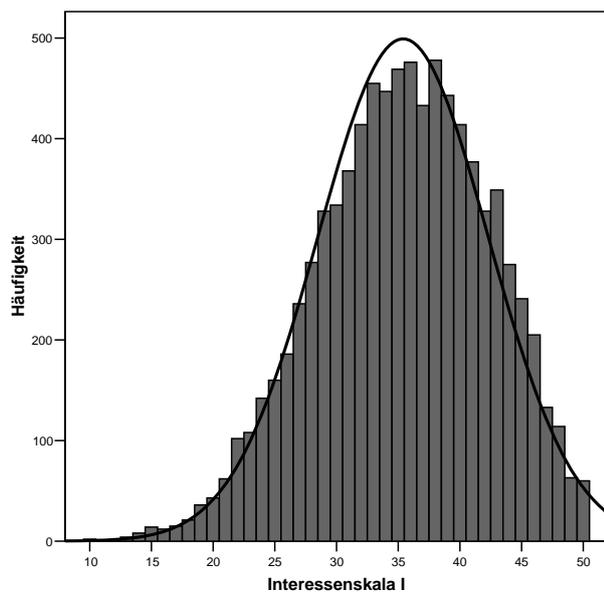


Abbildung 11.3: Scoreverteilung Skala INT-I

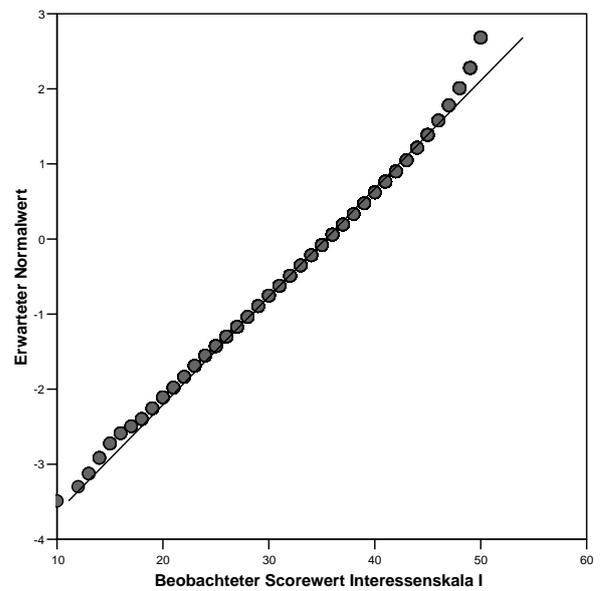


Abbildung 11.4: Q-Q-Diagramm Skala INT-I

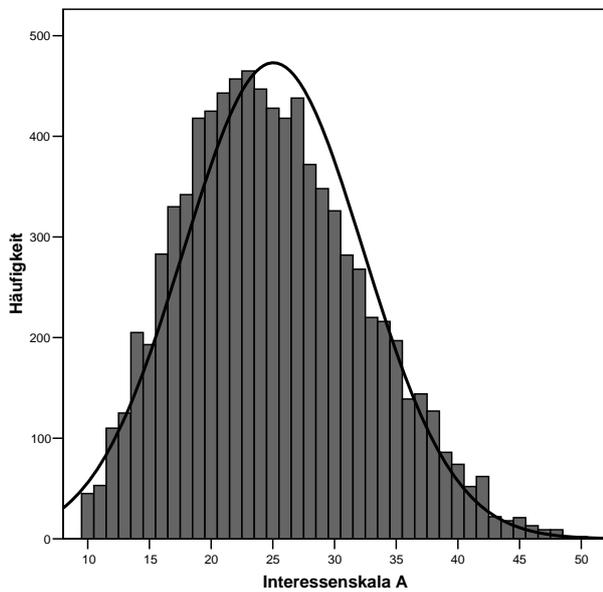


Abbildung 11.5: Scoreverteilung Skala INT-A

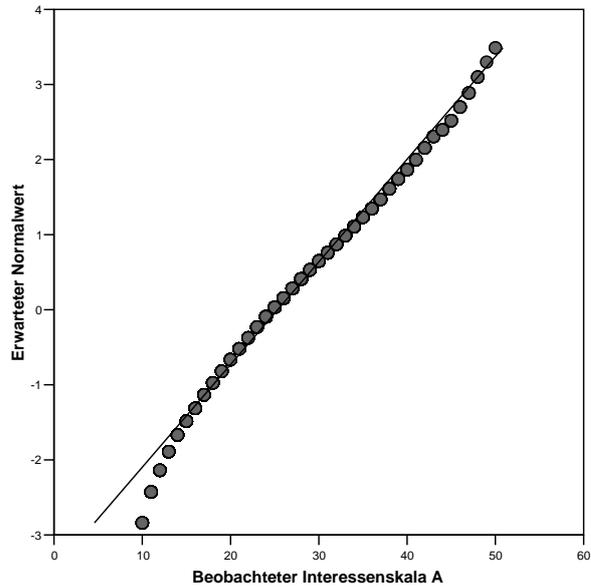


Abbildung 11.6: Q-Q-Diagramm Skala INT-A

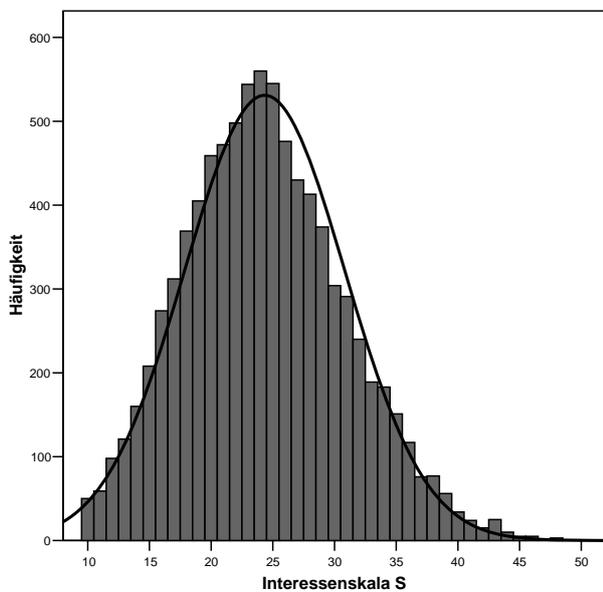


Abbildung 11.7: Scoreverteilung Skala INT-S

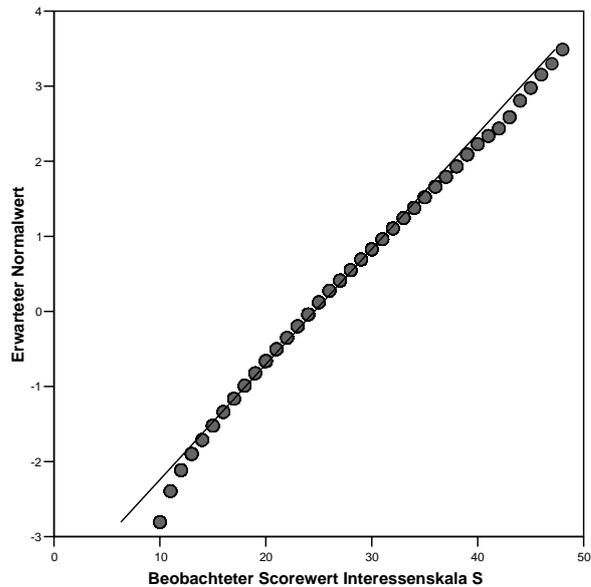


Abbildung 11.8: Q-Q-Diagramm Skala INT-S

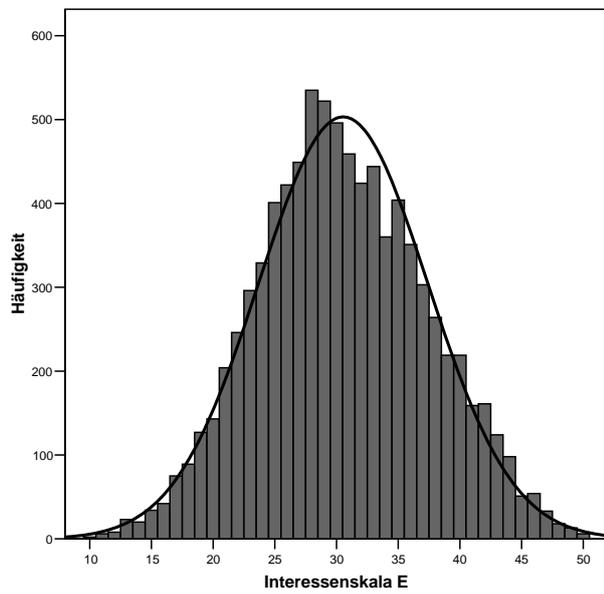


Abbildung 11.9: Scoreverteilung Skala INT-E

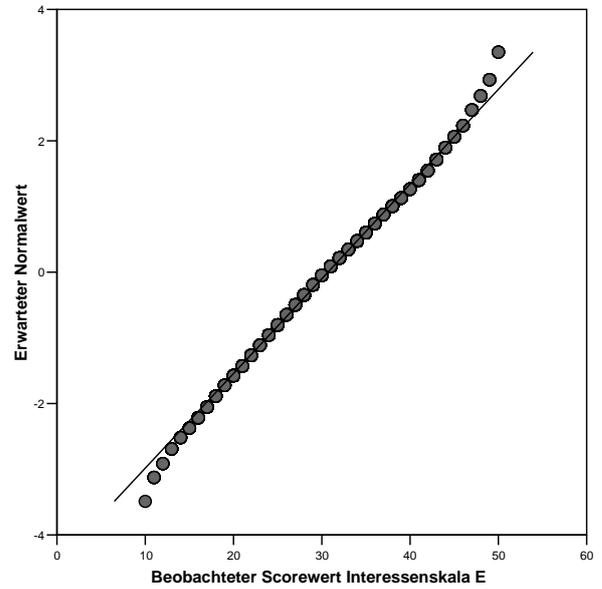


Abbildung 11.10: Q-Q-Diagramm Skala INT-E

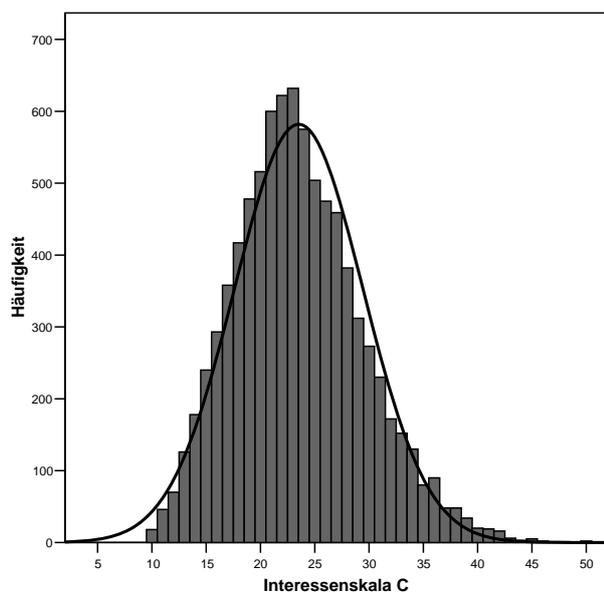


Abbildung 11.11: Scoreverteilung Skala INT-C

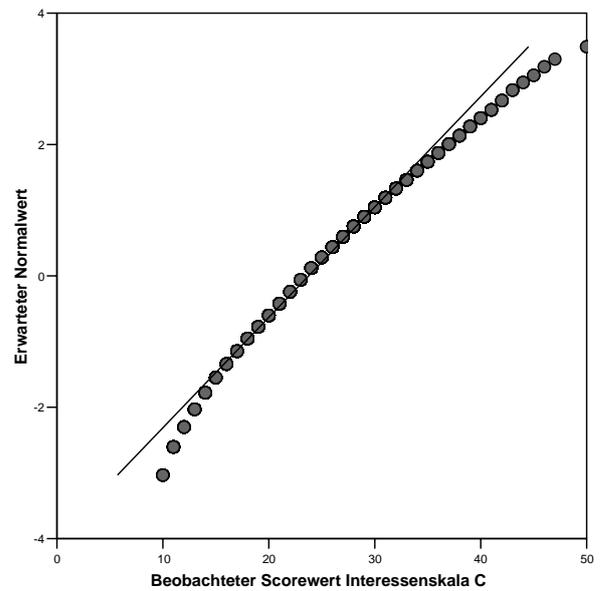


Abbildung 11.12: Q-Q-Diagramm Skala INT-C

Tabelle 11.1
Itemformulierungen INT - Teil 1

ID	Skala	Itemtext	Faktorladung	Skalen- mittelwert	Popularitäts- index	Trennschärfe
1	INT-C	Kosten für etwas kalkulieren	0,526	2,653	0,413	0,470
2	INT-R	Technische Geräte aufrüsten	0,678	3,263	0,566	0,594
3	INT-E	Diskussionen leiten	0,621	2,901	0,475	0,546
4	INT-A	Werke moderner Künstler beschreiben	0,578	1,801	0,200	0,524
5	INT-C	Abrechnungen erstellen	0,664	2,074	0,269	0,589
6	INT-I	Den Ursachen auf den Grund gehen	0,582	4,027	0,757	0,523
7	INT-S	Mit Jugendlichen arbeiten	0,604	2,774	0,443	0,533
8	INT-R	Maschinen in Betrieben errichten	0,794	2,715	0,429	0,727
9	INT-E	Arbeit von Personen koordinieren	0,672	3,346	0,586	0,588
10	INT-A	Wohnungen ideenreich modernisieren	0,543	2,976	0,494	0,426
11	INT-I	Neue Theorien entwickeln	0,728	3,612	0,653	0,660
12	INT-R	Technische Anlagen konzipieren	0,725	3,199	0,550	0,702
13	INT-E	Vereinbarungen aushandeln	0,710	2,879	0,470	0,637
14	INT-I	Sich hartnäckig in die Lösung eines Problems verbeißen	0,638	3,609	0,652	0,555
15	INT-S	Sich um Kranke kümmern	0,701	1,830	0,207	0,561
16	INT-E	Sprecher einer Gruppe sein	0,663	2,879	0,470	0,590
17	INT-A	Journalistisch arbeiten	0,529	2,478	0,369	0,511
18	INT-C	Anfragen bearbeiten	0,527	2,317	0,329	0,449
19	INT-S	Kinder erziehen	0,629	2,526	0,382	0,547
20	INT-R	Nachvollziehen, wie technische Anlagen funktionieren	0,690	3,601	0,650	0,684
21	INT-C	Etwas auf Vollständigkeit überprüfen	0,572	2,573	0,393	0,454
22	INT-I	Wissenschaftliche Ergebnisse analysieren	0,640	3,124	0,531	0,554
23	INT-S	Schülern Nachhilfe geben	0,554	2,520	0,380	0,479
24	INT-E	Jemanden von Standpunkten überzeugen	0,547	3,516	0,629	0,502
25	INT-A	Gedanken zeichnerisch festhalten	0,529	2,886	0,472	0,376
26	INT-C	Die Einhaltung von Regeln kontrollieren	0,557	2,377	0,344	0,472
27	INT-I	Formeln nachvollziehen	0,666	3,065	0,516	0,588
28	INT-S	Hilfesuchende beraten	0,605	2,816	0,454	0,540
29	INT-R	Werkstoffe auf Qualität prüfen	0,518	2,245	0,311	0,452
30	INT-A	Literatur verfassen	0,530	2,324	0,331	0,499

Alle Items korrelieren am höchsten mit dem erwarteten Faktor. Die Faktorladungen werden in der entsprechenden Spalte aufgeführt.

Tabelle 11.2
Itemformulierungen INT - Teil 1

ID	Skala	Itemtext	Faktorladungen		Trennschärfe	
			Skalen- mittelwert	Popularitäts- index		
31	INT-E	Projekte planen	0,597	3,647	0,662	0,519
32	INT-I	Neue Denkansätze entwickeln	0,700	3,658	0,665	0,645
33	INT-S	Schwerkranke auf ihrem letzten Weg begleiten	0,659	1,733	0,183	0,541
34	INT-R	Produktionsanlagen steuern	0,666	2,409	0,352	0,616
35	INT-E	Strategien für geschäftlichen Erfolg erlernen	0,628	3,215	0,554	0,537
36	INT-A	Fantasivolle Filme drehen	0,698	2,915	0,479	0,591
37	INT-C	Tabellen und Listen führen	0,752	2,301	0,325	0,629
38	INT-I	Aus Bekanntem Schlüsse ableiten	0,616	3,364	0,591	0,539
39	INT-S	Behinderte betreuen	0,692	1,704	0,176	0,568
40	INT-R	Maschinenbauteile herstellen	0,763	2,267	0,317	0,690
41	INT-E	Jemanden zur besseren Mitarbeit bewegen	0,516	2,760	0,440	0,510
42	INT-A	Logos und Symbole entwerfen	0,658	2,817	0,454	0,490
43	INT-C	Aufgaben nach klaren Ablaufplänen erledigen	0,602	2,678	0,419	0,476
44	INT-I	Sich in mathematische Sachverhalte hineinendenken	0,681	3,390	0,597	0,613
45	INT-S	Jemanden unterrichten	0,570	2,730	0,432	0,517
46	INT-R	Elektrische Schaltungen nachbauen	0,711	2,809	0,452	0,670
47	INT-A	Zeitkritische Theaterstücke schreiben	0,599	1,825	0,206	0,570
48	INT-C	Geschäftsbriefe verfassen	0,457	1,976	0,244	0,435
49	INT-S	Sich ehrenamtlich betätigen	0,536	2,628	0,407	0,460
50	INT-E	Löhne aushandeln	0,589	2,338	0,335	0,515
51	INT-A	Fotos für eine Zeitschrift machen	0,676	2,681	0,420	0,568
52	INT-S	Menschen in brenzligen Situationen helfen	0,589	3,080	0,520	0,535
53	INT-R	Konstruktionspläne aufertigen	0,646	2,867	0,467	0,575
54	INT-E	Vermarktung einer Idee ankurbeln	0,642	3,061	0,515	0,575
55	INT-A	Schauspielrollen übernehmen	0,519	2,346	0,337	0,496
56	INT-C	Materialien verwalten	0,611	2,020	0,255	0,520
57	INT-I	Auf einer abstrakten Ebene denken	0,703	3,531	0,633	0,622
58	INT-R	Fehler mit Messgeräten ausfindig machen	0,588	2,652	0,413	0,593
59	INT-C	Mit Bürosoftware arbeiten	0,587	2,548	0,387	0,496
60	INT-I	Sich mit Unerforschtem befassen	0,596	4,016	0,754	0,559

Alle Items korrelieren am höchsten mit dem erwarteten Faktor. Die Faktorladungen werden in der entsprechenden Spalte aufgeführt.

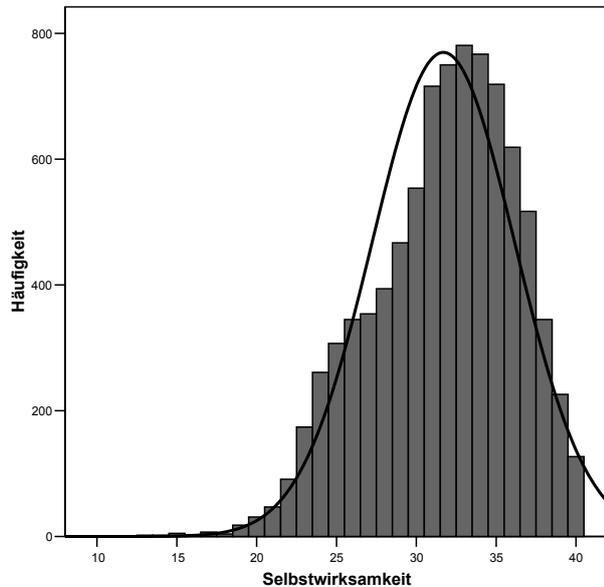


Abbildung 11.13: Scoreverteilung Skala SW

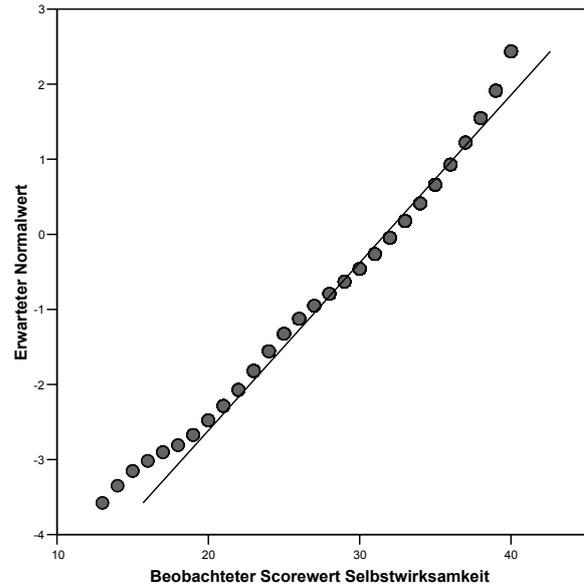


Abbildung 11.14: Q-Q-Diagramm Skala SW

Tabelle 11.3
Itemformulierungen SW

Nr.	Itemtext	Polung	Skalenmittelwert	Popularitätsindex	Trennschärfe
1	Auch wenn ich mich anstrenge, kann ich die Note Eins oder Zwei nicht erreichen.	u	3,254	0,751	0,535
2	Bei anstrengenden Klausuren weiß ich, wie ich den Lernstoff bewältigen kann.		3,035	0,678	0,377
3	Wenn ich mich gut vorbereite, erreiche ich mein Ziel.		3,425	0,808	0,386
4	Wenn etwas schwer wird, glaube ich kaum noch dran, dass ich eine Lösung finde.	u	2,980	0,660	0,606
5	Nach einer gewissen Zeit habe ich noch für alles eine Lösung gefunden.		3,116	0,705	0,367
6	Wenn ich etwas verstehen will, packe ich es auch.		3,444	0,815	0,392
7	Auch bei schweren Mathe-Aufgaben habe ich das Gefühl, sie lösen zu können.		3,043	0,681	0,399
8	Vieles packe ich aus Angst, es zu vermasseln, gar nicht erst an.	u	3,156	0,719	0,506
9	Auch bei guter Vorbereitung glaube ich nicht, dass ich erfolgreich sein werde.	u	3,312	0,771	0,569
10	Ich weiß genau, wie ich es anstellen muss, gute Noten zu schaffen.		2,930	0,643	0,360

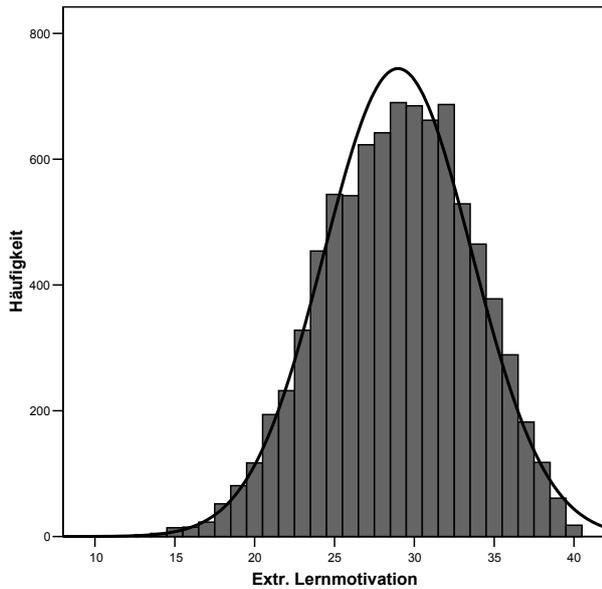


Abbildung 11.15: Scoreverteilung Skala ELM

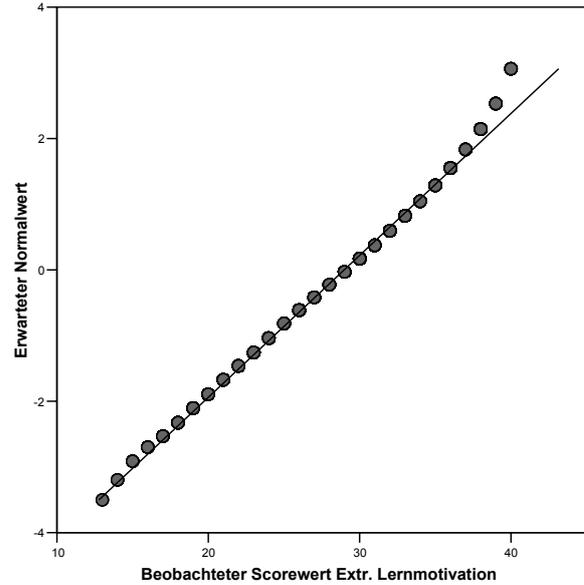


Abbildung 11.16: Q-Q-Diagramm Skala ELM

Tabelle 11.4
Itemformulierungen ELM

Nr.	Itemtext	Polung	Skalenmittelwert	Popularitätsindex	Trennschärfe
1	Es ist eine Niederlage für mich, wenn ich nicht die Note Eins oder Zwei habe.		2,416	0,472	0,377
2	Ich bin gerne besser als Andere.		3,288	0,763	0,244
3	Ich habe kaum konkrete Ziele im Leben, die ich auf jeden Fall erreichen will.	u	2,896	0,632	0,306
4	Mir ist es gleichgültig, ob ich in der Schule für etwas gelobt werde oder nicht.	u	2,952	0,651	0,363
5	Es ist unwichtig für mich, gute Noten zu haben.	u	3,258	0,753	0,499
6	Ich arbeite hart, weil ich mir hohe Ziele gesetzt habe.		2,819	0,606	0,463
7	Ich versuche, jedes Mal mehr zu erreichen als bisher.		3,025	0,675	0,354
8	Viele meinen, dass ich zu wenig für Prüfungen tue.	u	2,469	0,490	0,326
9	Mir ist es nicht wichtig, zu den Besten in der Klasse zu gehören.	u	2,624	0,541	0,471
10	Ich strenge mich an, um einen tollen Job zu bekommen.		3,229	0,743	0,398

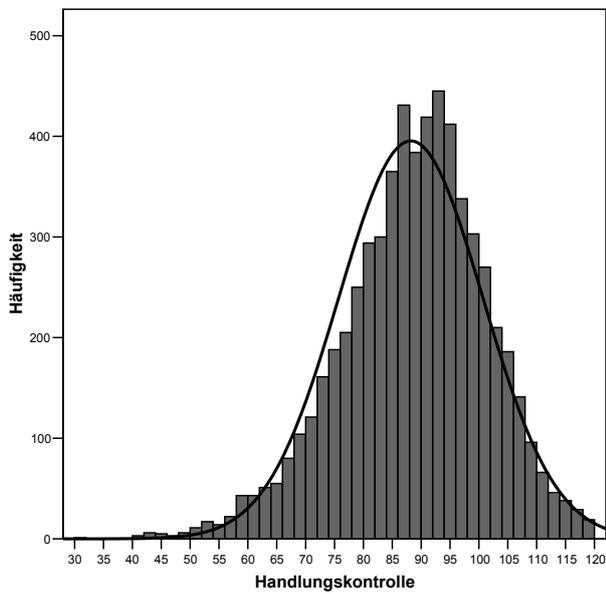


Abbildung 11.17: Scoreverteilung Skala HK

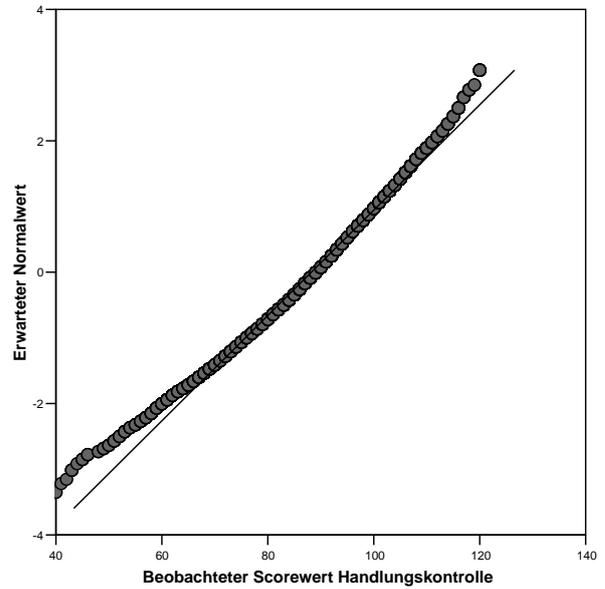


Abbildung 11.18: Q-Q-Diagramm Skala HK

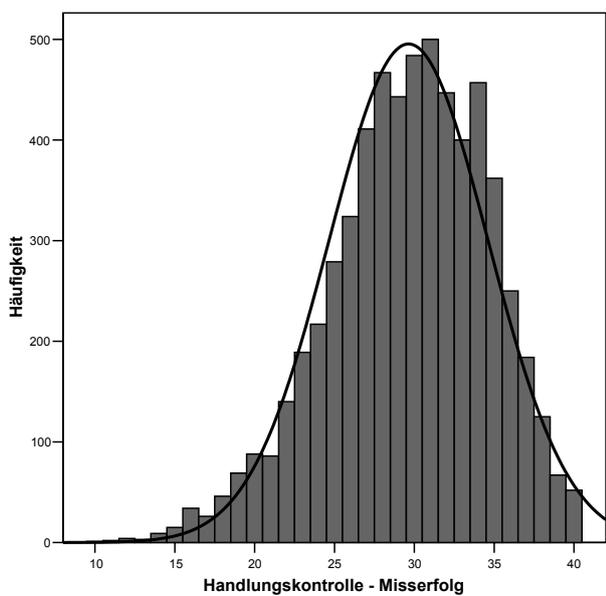


Abbildung 11.19: Scoreverteilung Skala HOM

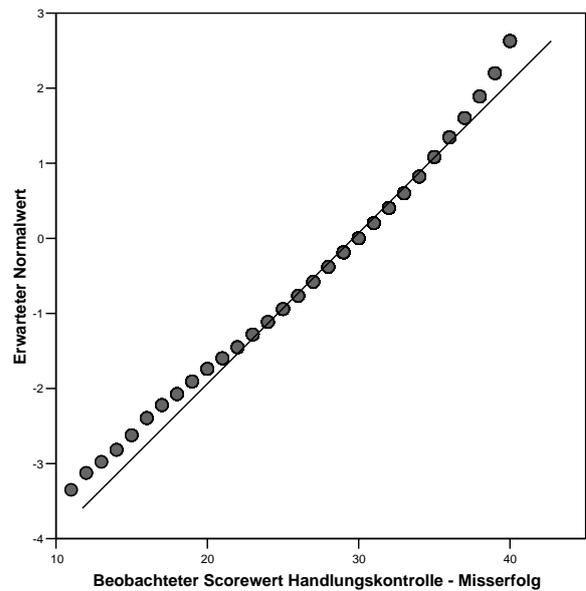


Abbildung 11.20: Q-Q-Diagramm Skala HOM

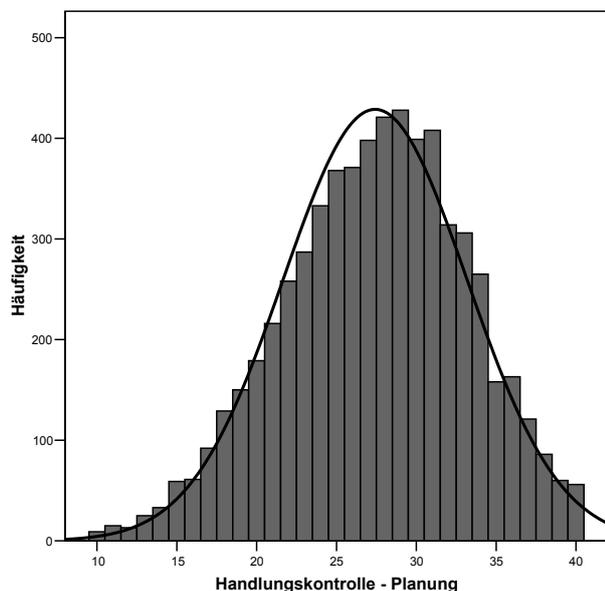


Abbildung 11.21: Scoreverteilung Skala HOP

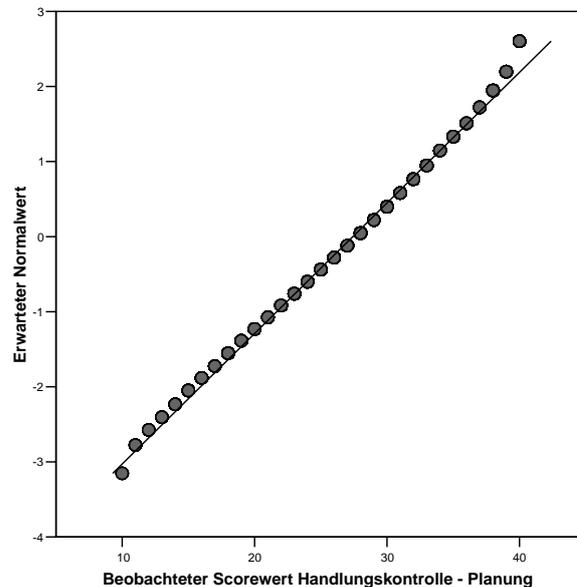


Abbildung 11.22: Q-Q-Diagramm Skala HOP

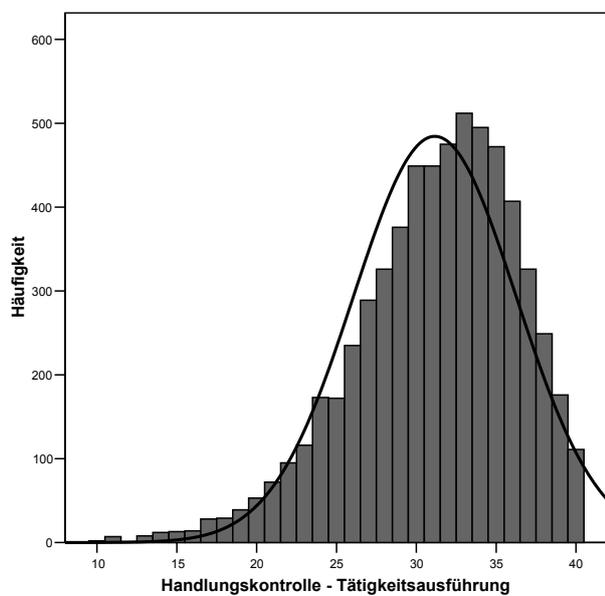


Abbildung 11.23: Scoreverteilung Skala HOT

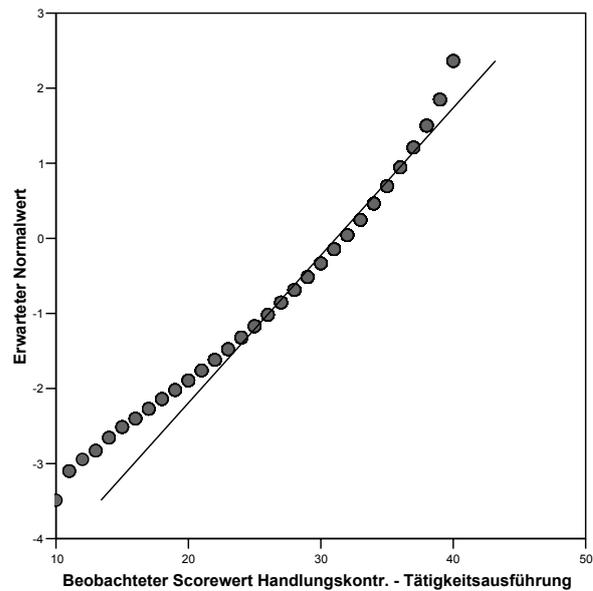


Abbildung 11.24: Q-Q-Diagramm Skala HOT

Tabelle 11.5
Itemformulierungen HK - Teil 1

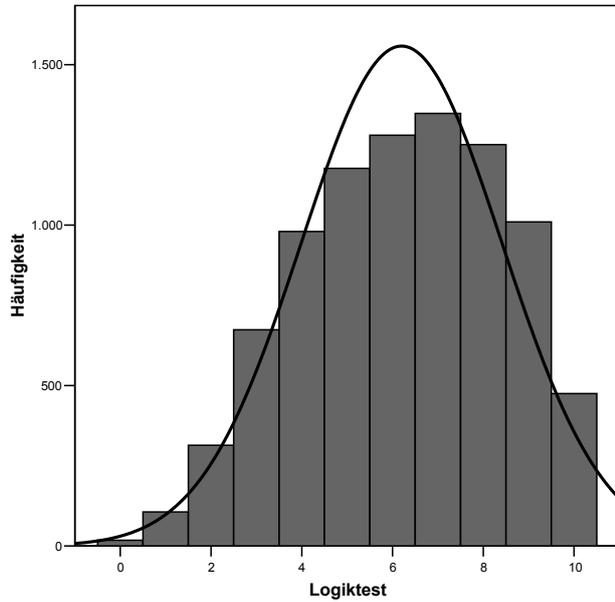
Nr.	Skala	Itemtext	Itemtext	Itemtext	Polung	Skalen- mittel- wert	Populäri- täts- index	Trenn- schärfe
1	HOT	Wenn ich etwas Interessan- tes lerne,	ertappe ich mich häufig da- bei, dass ich das Lernen nach kurzer Zeit unterbre- che.	lasse ich mich nicht ablen- ken.	u	3,207	0,736	0,428
2	HOM	Wenn ich in einer Klausur schlecht abgeschnitten ha- be,	hadere ich noch lange mit meinem Schicksal.	schiebe ich meinen Frust beiseite und versuche den Grund zu finden.	u	2,926	0,642	0,462
3	HOP	Wenn ich weiß, dass ich eine umfangreiche Hausar- beit fertig stellen muss,	denke ich oft daran, wie toll es doch wäre, wenn mir je- mand die Arbeit abnehmen würde.	überlege ich, wie ich die Ar- beit geschickt hinter mich bringe.		2,953	0,651	0,438
4	HOM	Wenn ein Lehrer meine Arbeitshaltung mangelhaft findet,	bin ich längere Zeit wie ge- lähmt.	bemühe ich mich sofort umso mehr.		2,965	0,655	0,352
5	HOP	Wenn ich lästige Formalitä- ten zu erledigen habe,	dauert es einige Zeit, bis ich mich dazu aufraffe.	fange ich die Sache so schnell wie möglich an.	u	2,153	0,384	0,522
6	HOT	Wenn ich gerade an etwas für mich Wichtigem arbei- te,	grübele ich häufig über ganz andere Dinge.	konzentriere ich mich sehr darauf.		3,387	0,796	0,434
7	HOP	Wenn ich mich auf eine Klausur vorbereiten muss,	beginne ich oft recht spät damit, weil ich nicht weiß, wie ich beginnen soll.	fange ich früh mit dem an, was mir am wichtigsten er- scheint.		2,454	0,485	0,533
8	HOT	Wenn ich mich zum Lernen hingesezt habe,	stehe ich auf, auch wenn ich der Lösung nicht näher ge- kommen bin.	bleibe ich bis zur Lösung am Ball.	u	2,885	0,628	0,491
9	HOM	Wenn ich etwas total ver- bocke,	lasse ich es besser beim nächsten Mal bleiben.	versuche ich es bald noch einmal.		3,141	0,714	0,373
10	HOT	Wenn ich bei einer Klausur sitze,	schweifen meine Gedanken schnell ab.	gelingt es mir, mich auf die Aufgaben zu konzentrieren.		3,430	0,810	0,382

Tabelle 11.6
Itemformulierungen HK - Teil 2

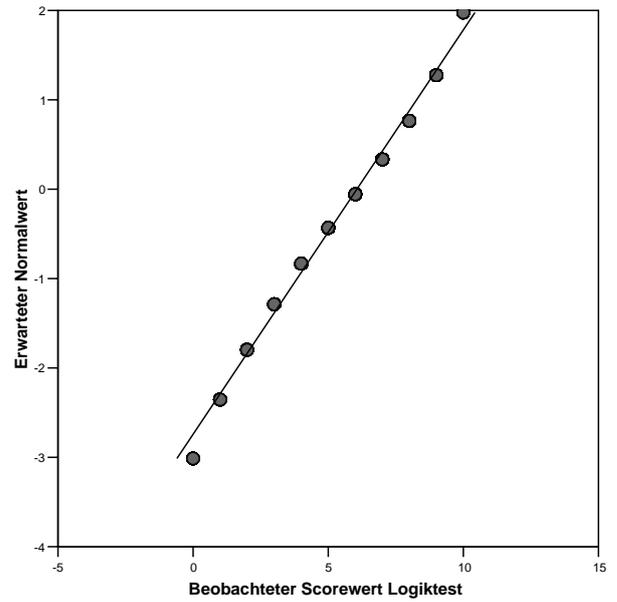
Nr.	Skala	Itemtext	Itemtext	Itemtext	Polung	Skalen- mittel- wert	Populari- täts- index	Trenn- schärfe
11	HOM	Wenn ich meinen ganzen Ehrgeiz daran gesetzt habe, eine bestimmte Klausur erfolgreich abzuschließen und es geht schief,	fällt es mir schwer, überhaupt noch etwas zu tun.	lasse ich den Kopf nicht hängen und fange mit den Vorbereitungen für die nächste Klausur an.	u	2,803	0,601	0,513
12	HOP	Wenn ich viele wichtige Arbeiten anzupacken habe,	weiß ich oft nicht, womit ich anfangen soll.	mache ich mir erst einmal einen neuen Arbeitsplan.		2,766	0,589	0,501
13	HOM	Wenn ich ein Schuljahr wiederholen müsste,	hätte ich ziemlich große Bedenken, dass ich auch dieses nicht schaffen würde.	wäre ich ziemlich optimistisch, weil ich den Stoff dann ja besser beherrschen würde.		3,337	0,779	0,324
14	HOT	Wenn ich Hausaufgaben mache,	plane ich dabei meine Freizeit.	löse ich erst einmal die Aufgaben.		3,055	0,685	0,444
15	HOM	Wenn mir in der Schule am selben Tag mehrere Dinge misslingen,	dann möchte ich am liebsten im Boden versinken.	dann bleibe ich tatkräftig, fast als wäre nichts passiert.	u	2,391	0,464	0,543
16	HOP	Wenn ich mir vorgenommen habe, mit etwas Wichtigem anzufangen,	kommt es häufig vor, dass ich meinen Plan im letzten Moment verwerfe.	führe ich mein Vorhaben auch durch.		2,867	0,622	0,528
17	HOM	Wenn ein Lehrer mich ungerecht behandelt hat,	kann mich das für mehrere Tage aus der Bahn werfen.	berührt mich das nicht lange.		2,848	0,616	0,475
18	HOP	Wenn ich geplant habe, schwierige Übungsaufgaben anzupacken,	gehen mir viele Dinge durch den Kopf, bevor ich mich richtig an die Aufgaben herammache.	überlege ich, welchen Weg ich zuerst ausprobiere.	u	2,948	0,649	0,401
19	HOT	Wenn ich mir vorgenommen habe, ein bestimmtes Pensum zu schaffen,	breche ich häufig trotzdem ab, ohne dieses zu erreichen.	höre ich nur sehr selten vorher auf.		2,930	0,643	0,509
20	HOP	Wenn ich viele wichtige Dinge planen muss,	schiebe ich die Planung länger vor mir her.	lege ich meist sofort los.		2,729	0,576	0,552

Tabelle 11.7
Itemformulierungen HK - Teil 3

Nr.	Skala	Itemtext	Itemtext	Itemtext	Polung	Skalen- mittel- wert	Populäri- täts- index	Trenn- schärfe
21	HOT	Wenn ich während einer Tätigkeit unterbrochen werde,	kommt es häufig vor, dass ich danach nicht mehr weitermache.	geht es direkt nach der Unterbrechung weiter.		2,711	0,570	0,429
22	HOT	Wenn ich mal so richtig büffeln muss,	kommt mir jede Unterbrechung gerade recht.	bleibe ich mit meinen Gedanken beim Thema.	u	2,891	0,630	0,492
23	HOM	Wenn ich mit einem Lehrer aneinander geraten bin,	beschäftigen mich die Wut und die Furcht vor Heimzahlung noch Tage.	kann ich die Wut und die Furcht vor Heimzahlung schnell vergessen.		2,890	0,630	0,485
24	HOP	Wenn ich mir vorgenommen habe, endlich mal aufgeschobene Aufgaben abzuarbeiten,	zögere ich, den Plan umzusetzen.	ziehe ich die Planung auch durch.	u	3,154	0,718	0,468
25	HOT	Wenn ich gerade eine schwere Matheaufgabe lösen möchte,	gehen mir alle möglichen Dinge durch den Kopf.	drehen sich meine Gedanken bis zur Lösung fast nur um die Aufgabe.		3,345	0,782	0,440
26	HOP	Wenn ich einen guten Vorschatz gefasst habe,	gelingt es mir eigentlich nie, den auch umzusetzen.	setze ich alles dran, den auch umzusetzen.		2,944	0,648	0,450
27	HOM	Wenn ich bei der Führerscheinprüfung durchgefallen bin,	bin ich für Tage nicht ansprechbar.	plane ich sofort die nächsten Schritte zum Führerschein.		3,301	0,767	0,444
28	HOP	Wenn ich mal etwas Unangenehmes anpacken muss,	kann es noch lange dauern, bis ich endlich damit anfangen kann.	mache ich es auch sofort.	u	2,449	0,483	0,488
29	HOT	Wenn gerade ich eine wichtige Sache mache,	unterbreche ich sie häufig, um mich anderen Dingen zu zuwenden.	ziehe ich sie auch durch.		3,316	0,772	0,539
30	HOM	Wenn ich bei einem Eignungstest schlecht abgeschnitten habe,	kann ich an nichts anderes mehr denken als an das Ergebnis.	kann ich mich trotzdem an schönen Dingen erfreuen.	u	3,036	0,679	0,409



(a)



(b)

Abbildung 11.25: Scoreverteilung und Q-Q-Diagramm Skala LOG

Keine Festplatte ist ein Prozessor.
Kein Arbeitsspeicher ist eine Festplatte.

Was kann man demnach über das Verhältnis von Arbeitsspeichern und Prozessoren **schlussfolgern**?

- a) Kein Arbeitsspeicher ist ein Prozessor.
- b) Einige Arbeitsspeicher sind Prozessoren.
- c) Keine der angegebenen Schlussfolgerungen ist korrekt.
- d) Alle Arbeitsspeicher sind Prozessoren.

Warte noch kurz bis du anklicken kannst! 2 / 11

p	0,726
r_{tt}	0,248
korrekte Lösung	c)

SelfAssessment

Abbildung 11.26: Logisches Schlussfolgern - Item 2; LOG4

Hier noch einmal ein bisschen kleiner...
 Ein Gebrauchtwagenhändler hat sechs nebeneinanderliegende Stellplätze in seinem Schauraum. Ein Audi, ein BMW, ein Golf, ein Opel, ein Mercedes und ein Porsche sollen von links nach rechts den Plätzen 1-6 zugeordnet werden. Aus Platzgründen und damit die einzelnen Fahrzeuge richtig zur Geltung kommen, müssen bei der Zuordnung der Autos zu den Stellplätzen folgende Bedingungen erfüllt sein:

- (1) Der Audi soll links vom BMW stehen.
- (2) Der Golf soll unmittelbar neben dem Porsche stehen.
- (3) Der Opel soll nicht unmittelbar neben dem Porsche stehen.
- (4) Der Mercedes soll an Position 4 stehen.
- (5) Der BMW soll an Position 5 stehen.

Welche der folgenden Zuordnungen **muss** demnach korrekt sein?

- a) Der Audi steht an Position 1.
- b) Der Golf steht an Position 1.
- c) Der Opel steht an Position 6.
- d) Der Porsche steht an Position 2.

Warte noch kurz bis du anklicken kannst! 3 / 11

p	0,697
r _{tt}	0,287
korrekte Lösung	c)

SelfAssessment

Abbildung 11.27: Logisches Schlussfolgern - Item3; LOG8

Die Bakterienstämme A-G weisen folgende Eigenschaften auf:

- (1) Alle A sind B.
- (2) Alle A sind C oder D.
- (3) Alle C und D sind A.
- (4) Alle E sind D.
- (5) Alle F sind G.
- (6) Kein G ist ein A.

Welcher **logische Schluss** lässt sich gemäß (1)-(6) ableiten?

- a) Kein F ist ein A.
- b) Kein G ist ein B.
- c) Kein G, das kein F ist, ist ein B.
- d) Einige G sind B.

Warte noch kurz bis du anklicken kannst! 4 / 11

p	0,528
r _{tt}	0,192
korrekte Lösung	a)

SelfAssessment

Abbildung 11.28: Logisches Schlussfolgern - Item 4; LOG9

Derselbe Text wieder ein bisschen verkleinert...

Für den Bau eines Generators werden zwei Bauteile der Gruppe 1 und drei Bauteile der Gruppe 2 benötigt. Innerhalb der Gruppe 1 stehen insgesamt drei verschiedene Ausführungen (A, B und C) zur Verfügung. Bei den Bauteilen der Gruppe 2 wird zwischen fünf verschiedenen Varianten (D, E, F, G und H) unterschieden. Aufgrund bestimmter Systemanforderungen und da nicht alle Komponenten miteinander kompatibel sind, gibt es folgende Einschränkungen, die bei der Konstruktion des neuen Generators zu beachten sind:

- (1) Entweder D oder G sollen für den Generator verbaut werden, aber nicht beide gemeinsam.
- (2) Die Bauteile A und C sollen nicht gemeinsam verbaut werden.
- (3) Entweder C oder E sollen für den Generator verbaut werden, aber nicht beide gemeinsam.
- (4) Entweder D oder F sollen für den Generator verbaut werden, aber nicht beide gemeinsam.
- (5) Kein Bauteil darf doppelt verbaut werden.

Wenn C verwendet wird, welches Bauteil darf dann **auf keinen Fall** ebenfalls verwendet werden?

a) B

b) D

c) F

d) G

Warte noch kurz bis du anklicken kannst! 5 / 11

p	0,747
r _{tt}	0,667
korrekte Lösung	b)

SelfAssessment

Abbildung 11.29: Logisches Schlussfolgern - Item 5; LOG10

In einer Gebäudewand sind übereinander sechs Kabel zu verlegen: Ein blaues, ein braunes, ein schwarzes, ein rotes, ein grünes und ein gelbes Kabel. Bei der Verlegung der Kabel müssen verschiedene Bedingungen eingehalten werden:

- (1) Das gelbe Kabel soll das dritte Kabel von oben sein.
- (2) Das schwarze Kabel soll entweder unmittelbar über oder unmittelbar unter dem roten Kabel liegen.
- (3) Das blaue Kabel soll unterhalb des braunen Kabels verlaufen.
- (4) Das grüne Kabel soll weder unmittelbar über dem roten noch unmittelbar unter dem roten Kabel verlegt werden.

Wenn das grüne Kabel das dritte von unten ist, welche der unten angegebenen Kabel verlaufen dann **auf jeden Fall unmittelbar übereinander**?

a) Blau und braun.

b) Blau und gelb.

c) Braun und grün.

d) Schwarz und grün.

Warte noch kurz bis du anklicken kannst! 6 / 11

p	0,620
r _{tt}	0,216
korrekte Lösung	a)

SelfAssessment

Abbildung 11.30: Logisches Schlussfolgern - Item 6; LOG12

Für einen Schaltkasten sind sechs Funktionsbereiche vorgesehen: Für die Bedienung einer Maschine, eines Roboters, eines Fließbandes, der Lichtanlage, der Klimaanlage und der automatischen Hallentore. Es wird eine technisch sinnvolle und benutzerfreundliche Anordnung der Funktionsbereiche in zwei Reihen zu je drei Schaltflächen gesucht. Dazu hat man einige Bedingungen aufgestellt:

- (1) Die Schaltfläche für das Fließband soll rechts von der Schaltfläche für die Maschine liegen.
- (2) Die Lichtschalter sollen links von der Schaltfläche für das Fließband liegen.
- (3) Die Schalter für die Hallentore sollen unmittelbar oberhalb der Lichtschalter angebracht werden.
- (4) Die Bedienfläche für den Roboter soll sich unmittelbar oberhalb der Schaltflächen für das Fließband befinden.

Welche der folgenden Aussagen muss entsprechend dieser vier Bedingungen **auf jeden Fall korrekt** sein?

- a) Die Schaltfläche zur Steuerung der Klimaanlage ist rechts von den Schaltern für das Hallentor.
- b) Die Schaltfläche zur Steuerung der Klimaanlage ist links von den Schaltern für das Hallentor.
- c) Die Lichtschalter befinden sich rechts von der Schaltfläche für die Maschinensteuerung.
- d) Die Schalter zur Bedienung des Fließbandes befinden sich rechts unten im Schaltkasten.

Warte noch kurz bis du anklicken kannst! 7 / 11

p	0,460
<i>r_{tt}</i>	0,254
korrekte Lösung	d)

SelfAssessment

Abbildung 11.31: Logisches Schlussfolgern - Item 7; LOG13

Bei einer Werksführung erklärt der Meister: "Wenn ich den roten Knopf drücke, dann bleibt die Maschine stehen."
Kurz später bleibt die Maschine stehen.

Was lässt sich daraus **schlussfolgern**?

- a) Der Meister hat den roten Knopf gedrückt.
- b) Der Meister hat den roten Knopf gedrückt oder er hat den roten Knopf nicht gedrückt.
- c) Der Meister hat den roten Knopf nicht gedrückt.
- d) Keine der angegebenen Schlussfolgerungen ist korrekt.

Warte noch kurz bis du anklicken kannst! 8 / 11

p	0,544
<i>r_{tt}</i>	0,291
korrekte Lösung	b)

SelfAssessment

Abbildung 11.32: Logisches Schlussfolgern - Item 8; LOG5

Ein Polizist notiert folgende Zeugenaussagen über einen Unfallhergang:

- (1) A passiert vor B.
- (2) A passiert vor C.
- (3) C passiert vor D.
- (4) E passiert vor A.
- (5) F passiert vor E.
- (6) G passiert vor F.

Welche der unten angeführten Überlegungen des Polizisten stellt **keinen folgerichtigen Schluss** dar?

- a) B passiert vor D.
- b) Von allen Ereignissen A-G passiert G als erstes.
- c) D passiert nach E.
- d) A passiert nach F.

Warte noch kurz bis du anklicken kannst! 9 / 11

SelfAssessment

p	0,528
r_{tt}	0,321
korrekte Lösung	a)

Abbildung 11.33: Logisches Schlussfolgern - Item 9; LOG3

Entweder der Kreis ist grün oder er ist nicht gepunktet.
Der Kreis ist gepunktet.

Welche der folgenden Aussagen stellt einen **logisch korrekten Schluss** dar?

- a) Der Kreis ist nicht grün.
- b) Der Kreis ist grün.
- c) Der Kreis ist gepunktet und nicht grün.
- d) Keine der angegebenen Schlussfolgerungen ist korrekt.

Warte noch kurz bis du anklicken kannst! 10 / 11

SelfAssessment

p	0,616
r_{tt}	0,233
korrekte Lösung	b)

Abbildung 11.34: Logisches Schlussfolgern - Item 10; LOG7

Einige Kutis sind Hanus.
Kein Gerla ist ein Kuti.

Was lässt sich daraus **schlussfolgern**?

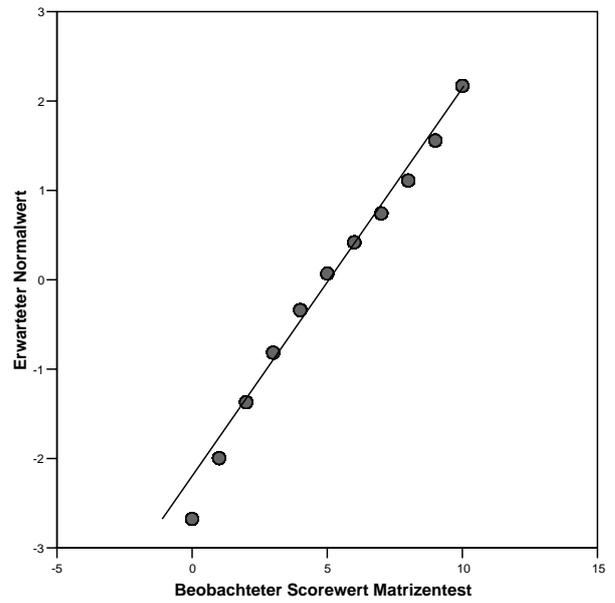
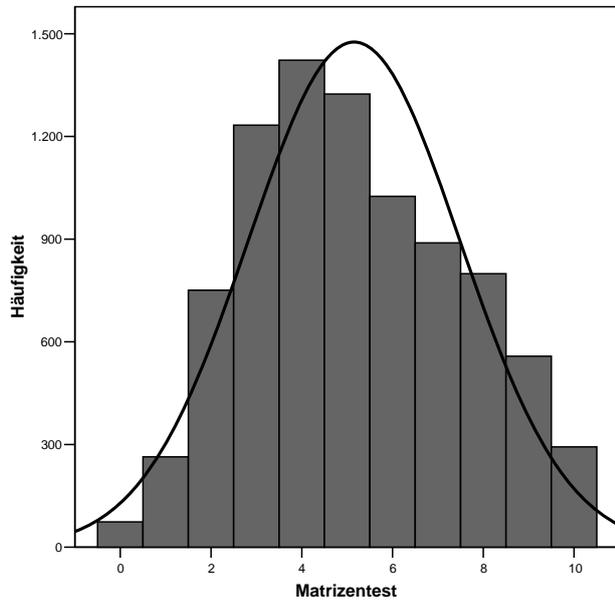
- a) Kein Hanu ist ein Gerla.
- b) Einige Hanus sind keine Gerlas.
- c) Einige Hanus sind Gerlas.
- d) Hanus sind entweder Kutis oder Gerlas.

Warte noch kurz bis du anklicken kannst! 11 / 11

SelfAssessment

p	0,528
<i>r_{tt}</i>	0,335
korrekte Lösung	b)

Abbildung 11.35: Logisches Schlussfolgern - Item 11; LOG6



(a) (b)
Abbildung 11.36: Scoreverteilung und Q-Q-Diagramm Skala MAT

Wart noch etwas bis du die Puzzlestücke ziehen kannst!
2 / 13

p	0,926
r_{tt}	0,211
korrekte Lösung	e)

SelfAssessment

Abbildung 11.37: Matrizentest - Item 2; MAT2

Wart noch etwas bis du die Puzzlestücke ziehen kannst! **3/13**

SelfAssessment

p	0,729
<i>r_{tt}</i>	0,283
korrekte Lösung	h)

Abbildung 11.38: Matrizentest - Item 3; MAT6

Wart noch etwas bis du die Puzzlestücke ziehen kannst! **4/13**

SelfAssessment

p	0,739
<i>r_{tt}</i>	0,233
korrekte Lösung	g)

Abbildung 11.39: Matrizentest - Item 4; MAT8

Wart noch etwas bis du die Puzzlestücke ziehen kannst! 5/13

SelfAssessment

Abbildung 11.40: Matrizentest - Item 5; MAT10

p	0,441
<i>r_{tt}</i>	0,407
korrekte Lösung	d)

Wart noch etwas bis du die Puzzlestücke ziehen kannst! 6/13

SelfAssessment

Abbildung 11.41: Matrizentest - Item 6; MAT11

p	0,465
<i>r_{tt}</i>	0,243
korrekte Lösung	e)

Wart noch etwas bis du die Puzzlestücke ziehen kannst! **7/13**

SelfAssessment

p	0,330
r_{tt}	0,258
korrekte Lösung	d)

Abbildung 11.42: Matrizentest - Item 7; MAT13

Wart noch etwas bis du die Puzzlestücke ziehen kannst! **8/13**

SelfAssessment

p	0,381
r_{tt}	0,411
korrekte Lösung	h)

Abbildung 11.43: Matrizentest - Item 8; MAT15

Wart noch etwas bis du die Puzzlestücke ziehen kannst! 9/13

SelfAssessment

p	0,456
<i>r_{tt}</i>	0,413
korrekte Lösung	h)

Abbildung 11.44: Matrizentest - Item 9; MAT5

Wart noch etwas bis du die Puzzlestücke ziehen kannst! 10/13

SelfAssessment

p	0,373
<i>r_{tt}</i>	0,406
korrekte Lösung	b)

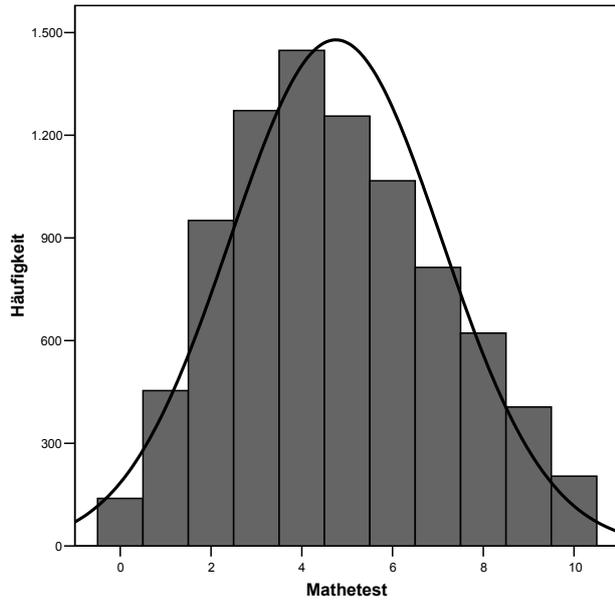
Abbildung 11.45: Matrizentest - Item 10; MAT12

Wart noch etwas bis du die Puzzlestücke ziehen kannst! 11 / 13

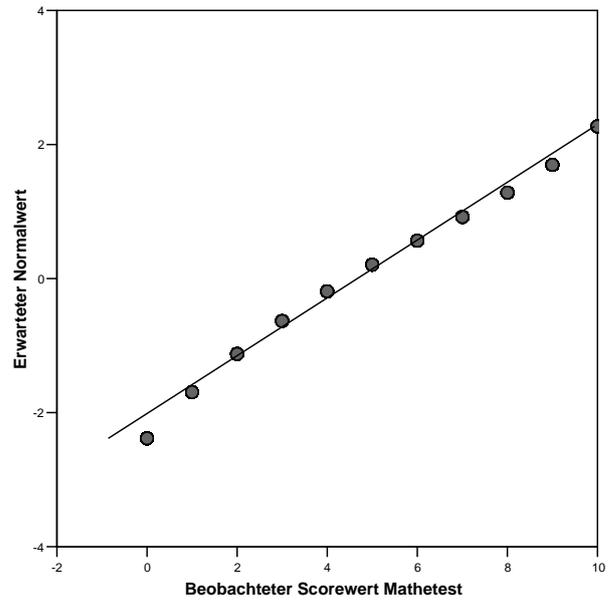
SelfAssessment

p	0,314
<i>r_{tt}</i>	0,538
korrekte Lösung	d)

Abbildung 11.46: Matrizentest - Item 11; MAT7



(a)



(b)

Abbildung 11.47: Scoreverteilung und Q-Q-Diagramm MATH

Der Preis P_a für einen CD-Brenner im Computerladen Hard & Soft liegt um x % unter dem Preis P_b , den ein Computergroßmarkt verlangt. Wie hoch ist der Preis P_a bei Hard & Soft?

a) Keine der angegebenen Lösungen ist korrekt.

b)
$$P_a = P_b - \frac{100}{P_b \cdot x}$$

c)
$$P_a = P_b - \frac{P_b \cdot x}{100}$$

d)
$$P_a = P_b - \frac{100 \cdot x}{P_b}$$

Warte noch kurz bis du anklicken kannst! 2 / 11

p	0,790
r_{tt}	0,229
korrekte Lösung	c)

SelfAssessment

Abbildung 11.48: Mathematische Fähigkeiten - Item 2; MATH5

$$\frac{k^6 \cdot n^3 - m^3 \cdot l^3}{m^3 \cdot n^3} + \sqrt[5]{o} \cdot \sqrt[5]{p} = ?$$

a) $= \left(\frac{k^2}{m}\right)^3 - \left(\frac{l}{n}\right)^3 + \sqrt[5]{o \cdot p}$

b) Keine der angegebenen Lösungen ist korrekt.

c) $= \left(\frac{k}{m}\right)^3 - \left(\frac{l}{n}\right)^3 + \sqrt[10]{o \cdot p}$

d) $= k^2 - l^3 + \sqrt[5]{o \cdot p}$

Warte noch kurz bis du anklicken kannst! 3 / 11

SelfAssessment

p	0,634
r _{tt}	0,354
korrekte Lösung	a)

Abbildung 11.49: Mathematische Fähigkeiten - Item 3; MATH8

$$\frac{\frac{m}{n} + \frac{n}{m}}{3} = ?$$

a) $= \frac{m^2 + n^2}{3}$

b) Keine der angegebenen Lösungen ist korrekt.

c) $= \frac{3 \cdot (m+n)^2}{m \cdot n}$

d) $= \frac{m^2 \cdot n^2}{3}$

Warte noch kurz bis du anklicken kannst! 4 / 11

SelfAssessment

p	0,555
r _{tt}	0,348
korrekte Lösung	a)

Abbildung 11.50: Mathematische Fähigkeiten - Item 4; MATH2

$(-x^3 \cdot y^2 + x \cdot z \cdot a) \div (x^2 \cdot y^3) = ?$

a)	Keine der angegebenen Lösungen ist korrekt.
b)	$= -\frac{x}{y} + x^{-1} \cdot z \cdot a \cdot y^{-3}$
c)	$= -\left(\frac{y}{x}\right)^{-1} + \frac{x \cdot y^3}{z \cdot a}$
d)	$= -x \cdot y^{-1} + \frac{x^3 \cdot z \cdot a}{y^3}$

Warte noch kurz bis du anklicken kannst! 5 / 11

SelfAssessment

p	0,507
r _{tt}	0,396
korrekte Lösung	b)

Abbildung 11.51: Mathematische Fähigkeiten - Item 5; MATH10

In einem Textilunternehmen werden 40 m Stoff pro Stunde zu Hosen verarbeitet. Ein zweites, moderneres Unternehmen, das auch Hosen produziert, verarbeitet 70 m Stoff pro Stunde. Die Produktion im zweiten Unternehmen beginnt morgens eine halbe Stunde später als im ersten Unternehmen. Wenn wir den Zeitpunkt betrachten, an dem beide Unternehmen gleich viel Stoff verbraucht haben, wie viele Meter Stoff wurden dann **jeweils** schon zu Hosen verarbeitet?

a)	$\frac{245}{3} \approx 81,67m$
b)	70m
c)	Keine der angegebenen Lösungen ist korrekt.
d)	$\frac{140}{3} \approx 46,67m$

Warte noch kurz bis du anklicken kannst! 6 / 11

SelfAssessment

p	0,467
r _{tt}	0,275
korrekte Lösung	d)

Abbildung 11.52: Mathematische Fähigkeiten - Item 6; MATH15

$(x - 1) \cdot (x - 2) \cdot (x - 3) > 0$

a)	Diese Ungleichung ist unerfüllbar.
b)	Keine der angegebenen Lösungen ist korrekt.
c)	$x > 3 \vee x < 1$
d)	$x > 3 \vee 1 < x < 2$

Warte noch kurz bis du anklicken kannst! 7 / 11

p	0,383
r _{tt}	0,326
korrekte Lösung	d)

SelfAssessment

Abbildung 11.53: Mathematische Fähigkeiten - Item 7; MATH3

Um ein Staubecken mit Wasser zu füllen, braucht eine Pumpe sechs Stunden. Eine andere Pumpe braucht dafür vier Stunden. Durch ein Abflussrohr kann das gesamte Staubecken in einer Zeit von zwölf Stunden geleert werden. Wenn beide Pumpen gleichzeitig arbeiten und auch das Abflussrohr geöffnet ist, **in welcher Zeit t ist dann das anfangs leere Becken voll?**

a)	Keine der angegebenen Lösungen ist korrekt.
b)	t = 5 Stunden
c)	t = 3 Stunden
d)	t = 4 Stunden

Warte noch kurz bis du anklicken kannst! 8 / 11

p	0,427
r _{tt}	0,287
korrekte Lösung	c)

SelfAssessment

Abbildung 11.54: Mathematische Fähigkeiten - Item 8; MATH14

Eine Firma ist von drei Teilhabern mit den folgenden Vermögensanteilen gegründet worden: r Euro, s Euro und t Euro. Der Reingewinn der Firma beläuft sich auf einen Betrag, der 60% von t entspricht. Nach welcher der folgenden Formeln lässt sich berechnen, wie viel (x Euro) vom Reingewinn dem Partner mit der Einlage r zusteht, wenn jeder Partner einen Anteil entsprechend seiner Investition erhält?

a) $x = \frac{0,6 \cdot r \cdot t}{r + s + t}$

b) $x = \frac{r + s + t}{0,6 \cdot t} \cdot r$

c) $x = \frac{r \cdot t}{0,6 \cdot (r + s + t)}$

d) Keine der angegebenen Lösungen ist korrekt.

Warte noch kurz bis du anklicken kannst! 9 / 11

SelfAssessment

Abbildung 11.55: Mathematische Fähigkeiten - Item 9; MATH6

p	0,365
r _{tt}	0,351
korrekte Lösung	a)

Für Ausschachtungsarbeiten setzt ein Bauunternehmer einen Bagger ein, der die Arbeit in 12 Tagen bewältigen könnte. Um aber den Prozess zu beschleunigen, wird nach 2 Tagen zusätzlich ein zweiter Bagger eingesetzt, der allerdings nur ein Leistungsvermögen von 80 % gegenüber dem ersten Bagger hat. In wie vielen Tagen (x) sind die **verbleibenden Ausschachtungsarbeiten** abgeschlossen?

a) $x = \frac{50}{9} \approx 5,56$

b) $x = \frac{9}{2} = 4,5$

c) Keine der angegebenen Lösungen ist korrekt.

d) $x = \frac{20}{3} \approx 6,67$

Warte noch kurz bis du anklicken kannst! 10 / 11

SelfAssessment

Abbildung 11.56: Mathematische Fähigkeiten - Item 10; MATH12

p	0,333
r _{tt}	0,254
korrekte Lösung	a)

Für Ausschachtungsarbeiten setzt ein Bauunternehmer einen Bagger ein, der die Arbeit in 12 Tagen bewältigen könnte. Um aber den Prozess zu beschleunigen, wird nach 2 Tagen zusätzlich ein zweiter Bagger eingesetzt, der allerdings nur ein Leistungsvermögen von 80 % gegenüber dem ersten Bagger hat. In wie vielen Tagen (x) sind die **verbleibenden Ausschachtungsarbeiten** abgeschlossen?

a)	$x = \frac{50}{9} \approx 5,56$
b)	$x = \frac{9}{2} = 4,5$
c)	Keine der angegebenen Lösungen ist korrekt.
d)	$x = \frac{20}{3} \approx 6,67$

Warte noch kurz bis du anklicken kannst! 10 / 11

p	0,295
r_{tt}	0,246
korrekte Lösung	d)

SelfAssessment

Abbildung 11.57: Mathematische Fähigkeiten - Item 11; MATH4

Curriculum Vitae

Lebenslauf des Autors

Name: Alexander Zimmerhofer

geboren am: 07.12.1976 in Duisburg

Ausbildung

08.1983–08.1987 Grundschule St. Lucia, Geldern-Walbeck

08.1987–06.1996 Friedrich-Spee-Gymnasium, Geldern

07.1996–05.1997 Diakonisches Werk, Geldern (Zivildienst)

10.1997–09.1999 Heinrich-Heine Universität, Düsseldorf (Grundstudium der Psychologie - Diplom)

10.1999–01.2003 Rheinisch-Westfälische Technische Hochschule Aachen (Hauptstudium der Psychologie - Diplom)

01.2003 Abgabe der Diplomarbeit und Abschluss des Studiums

Berufliche Tätigkeiten

08.2003–05.2006 Wissenschaftliche Hilfskraft bzw. Mitarbeiter am Lehrstuhl für Betriebs- und Organisationspsychologie der RWTH Aachen

seit 06.2006 Human-Resources-Berater bei der ITB Consulting GmbH, Bonn

Düsseldorf, 11. Juni 2008