

Die relative Bedeutung kognitiver und non-kognitiver Bedingungsfaktoren bei der Vorhersage des Schulerfolgs

Vergleich der prädiktiven Validität von
Arbeitsgedächtnis, Intelligenz, Motivation und SES

Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Philosophie
der Philosophischen Fakultäten
der Universität des Saarlandes

vorgelegt von
Heike Sabine Weber
aus Neunkirchen

Saarbrücken 2012

Dekan:

Prof. Dr. Jochen Kubiniok

Berichterstatter/innen:

Prof. Dr. Frank M. Spinath

Prof. Dr. Gisa Aschersleben

Tag der Disputation: 12.06.2012

Danksagung

Das Gelingen dieser Arbeit war an die Unterstützung vieler Personen gebunden, bei denen ich mich an dieser Stelle besonders bedanken möchte.

So danke ich Prof. Dr. Frank M. Spinath für die Ermöglichung dieser Studie, sowie für die Betreuung während der letzten drei Jahre. Außerdem möchte ich mich bei allen Mitarbeitern/innen der AE Differentielle Psychologie bedanken, für die gute Zeit und die großartige Unterstützung. Ein großer Dank gilt auch meinen HiWis Marina Zwipf, die durch ihren Einsatz die Datenerhebung maßgeblich unterstützt hat und Lisa Erhardt, die mir trotz meines Aufenthalts in China stets zuverlässig zugearbeitet hat. Selbstverständlich wäre diese Studie nicht möglich gewesen, ohne das große Engagement der beteiligten Lehrkräfte, sowie der Kinder und Eltern, die sich zur Teilnahme bereit erklärt haben. Ihnen möchte ich ebenfalls herzlich danken, ebenso wie dem Ministerium für Bildung für die Genehmigung der Studie.

Ein weiterer Dank gilt den zahlreichen fleißigen Korrekturlesern/innen dieser Arbeit, deren Hilfe mir eine sehr große Unterstützung war, ebenso wie meinen Freunden und Kollegen, die mir in den schwierigsten Phasen stets Mut gemacht und jegliche Art von Unterstützung und Aufmunterung geboten haben. Danke Juliana, Conny, Luzie, Martina, André, Tobias, Maja, Nicolas, Simone, Kathrin, Katrin und Ina!

Mein Dank gilt in besonderem Maße meiner Mutter für ihre Stärke, sowie meiner gesamten Familie und Raschid Dörr, die mir stets zur Seite standen und mich mit unerschöpflicher Geduld durch die anstrengendsten Wochen begleitet haben. Raschid, ich wäre verhungert ohne dich!

Saarbrücken, 16.01.2012

Inhaltsverzeichnis

Abbildungsverzeichnis.....	VII
Tabellenverzeichnis.....	IX
Abkürzungsverzeichnis.....	XI
Einleitung	1
I Theoretische und empirische Grundlagen.....	7
1.1 Arbeitsgedächtnis.....	7
1.1.1 Einführung.....	8
1.1.2 Modellvorstellungen	9
1.1.3 Empirische Befunde zur Frage der Domänen-Generalität bzw. -Spezifität des Arbeitsgedächtnisses.....	16
1.1.4 Entwicklung des Arbeitsgedächtnisses im Kindesalter	18
1.1.5 Zusammenfassende Betrachtung.....	22
1.2 Intelligenz.....	24
1.2.1 Überblick über das Themengebiet der Intelligenzforschung	24
1.2.2 Zum Begriff der allgemeinen Intelligenz.....	26
1.2.3 Strukturmodelle	27
1.2.4 Zur Bedeutung des <i>g</i> -Faktors.....	30
1.2.5 Entwicklung und Stabilität.....	32
1.2.6 Zur Beziehung zwischen WM und Intelligenz	34
1.3 Motivation	39
1.3.1 Konzeptionen	41
1.3.2 Entwicklung und Stabilität der Erwartungs- und Wertkomponenten.....	51
1.3.3 Zusammenfassung.....	56
1.4 Sozioökonomischer Status.....	57
1.5 Schulerfolg	62
1.5.1 Kognitive Prädiktoren und Schulerfolg.....	64
1.5.2 Non-kognitive Prädiktoren und Schulerfolg.....	71
1.5.3 Sozioökonomischer Status und Schulerfolg.....	79
II Fragestellung und Hypothesen.....	82
2.1 Die Bedeutung der kognitiven Prädiktoren WM und gf für die Vorhersage des Schulerfolgs	83
2.2 Die relative Bedeutung kognitiver und non-kognitiven Prädiktoren bei der Vorhersage des Schulerfolgs	84

2.3	Die Bedeutung des SES für die Vorhersage des Schulerfolgs.....	86
III	Material und Methoden	87
3.1	Graduiertenkolleg IRTG	87
3.2	Planung und Durchführung der Datenerhebung	88
3.3	Beschreibung der Stichprobe	89
3.4	Erhebungsinstrumente	95
3.4.1	Erfassung der kognitiven Fähigkeiten.....	95
3.4.2	Erfassung der non-kognitiven Faktoren	101
3.4.3	Erfassung des sozioökonomischen Status	102
3.4.4	Erfassung der Schulnoten	102
3.5	Analysemethoden	103
3.5.1	Kommunalitätenanalyse	103
3.5.2	Strukturgleichungsmodellierung	104
IV	Ergebnisse	118
4.1	Deskriptive Statistiken und korrelative Zusammenhänge	118
4.1.1	Deskriptive Statistiken und Binnenstruktur der kognitiven Prädiktoren.....	118
4.1.2	Deskriptive Statistiken und Binnenstruktur der non-kognitiven Prädiktoren.....	123
4.1.3	Deskriptive Statistiken und Binnenstruktur des SES	125
4.1.4	Korrelative Betrachtung der Zusammenhänge zwischen den Prädiktoren.....	127
4.1.5	Deskriptive Statistiken und Binnenstruktur des Kriteriums Schulerfolg	128
4.1.6	Korrelative Betrachtung der Prädiktor-Kriteriums- Zusammenhänge	131
4.2	Die Bedeutung der kognitiven Faktoren WM und gf als Prädiktoren des Schulerfolgs	133
4.2.1	Regressionsanalytische Befunde	133
4.2.2	Strukturgleichungsanalytische Befunde.....	137
4.3	Relative Bedeutung kognitiver und non-kognitiver Faktoren als Prädiktoren des Schulerfolgs	154
4.3.1	Prüfung des Messmodells der non-kognitiven Faktoren	154
4.3.2	Prognosemodell der kognitiven und non-kognitiven Faktoren und des SES als Prädiktoren des Schulerfolgs.....	157

4.3.3	Zwei-Ebenen-Modellierung der kognitiven und non-kognitiven Faktoren und des SES als Prädiktoren des Schulerfolgs	169
4.3.4	Zusammenfassung.....	175
V	Diskussion.....	177
5.1	Erhebung und Binnenstruktur der kognitiven Fähigkeiten.....	177
5.1.1	Arbeitsgedächtnis.....	177
5.1.2	Intelligenz.....	179
5.2	Erhebung und Binnenstruktur der non-kognitiven Faktoren und des SES.....	180
5.3	Erhebung des Kriteriums Schulerfolg	183
5.4	Die Bedeutung der kognitiven Faktoren WM und gf als Prädiktoren des Schulerfolgs	184
5.4.1	Zum Zusammenhang von WM und gf.....	185
5.4.2	Vergleich der prädiktiven Validität von WM und gf auf manifester Ebene	186
5.4.3	Vergleich der prädiktiven Validität von WM und gf auf latenter Ebene.....	188
5.5	Die relative Bedeutung kognitiver und non-kognitiver Faktoren.....	192
5.5.1	Die relative Bedeutung kognitiver und non-kognitiver Prädiktoren in Deutsch	192
5.5.2	Die relative Bedeutung kognitiver und non-kognitiver Prädiktoren in Mathematik.....	197
5.6	Zur relativen Bedeutung des SES als Prädiktor des Schulerfolgs	200
5.7	Fazit.....	205
5.8	Einschränkungen und Grenzen der Studie.....	206
5.9	Ausblick	210
5.10	Praktische Implikationen der Relevanz non-kognitiver Bedingungsfaktoren	213
	Literaturverzeichnis.....	215
	Anhang A Tabellenanhang	247
	Anhang B Materialanhang	252
	Anhang C Lebenslauf	258

Abbildungsverzeichnis

Abbildung 1: Erwartungs-Wert-Modell der Leistungsmotivation (Eccles & Wigfield, 2002).....	48
Abbildung 2: Modell der Bedingungsfaktoren schulischer Leistung (Helmke & Schrade, 2006).....	63
Abbildung 3: Verteilung der Schulabschlüsse innerhalb der Stichprobe....	93
Abbildung 4: Verteilung der Schulabschlüsse der Erhebungsstichprobe im Vergleich zur DESTATIS-Stichprobe	94
Abbildung 5: Darstellung von Mess- und Strukturmodell der Strukturgleichungsanalyse (Weiber & Mülhhaus, 2010)	111
Abbildung 6: Häufigkeitsverteilung der Schulnoten getrennt nach Schuljahr und -fach (Prozentwerte).....	130
Abbildung 7: Spezifische und gemeinsame Varianzanteile von gf und WM im Fach Deutsch.....	135
Abbildung 8: Spezifische und gemeinsame Varianzanteile von gf und WM im Fach Mathematik	137
Abbildung 9: Zweifaktorielles Messmodell der kognitiven Fähigkeiten	138
Abbildung 10: Einfaktorielles Messmodell der kognitiven Fähigkeiten.....	139
Abbildung 11: Schematische Darstellung der schrittweisen Fixierung nicht signifikanter Pfade in Deutsch	141
Abbildung 12: Parsimonischstes Prognosemodell der kognitiven Faktoren und des SES in Deutsch (Modell 1a).....	144
Abbildung 13: Parsimonischstes Prognosemodell der kognitiven Faktoren und des SES in Mathematik (Modell 2a)	147
Abbildung 14: Zwei-Ebenen-Modell der kognitiven Fähigkeiten und des SES	149
Abbildung 15: Parsimonischstes Zwei-Ebenen-Modell der kognitiven Faktoren und des SES in Deutsch und Mathematik (Modell 3 und Modell 4a).....	150
Abbildung 16: Zweifaktorielles Messmodell der non-kognitiven Faktoren	155
Abbildung 17: Einfaktorielles Messmodell der non-kognitiven Faktoren ..	156
Abbildung 18: Prognosemodell der kognitiven und non-kognitiven Faktoren und des SES.....	158

Abbildung 19: Schematische Darstellung der schrittweisen Fixierung nicht signifikanter Pfade in Deutsch	160
Abbildung 20: Schematische Darstellung der schrittweisen Fixierung nicht signifikanter Pfade in Mathematik	165
Abbildung 21: Parsimonischstes Prognosemodell der kognitiven und non-kognitiven Faktoren, sowie des SES in Mathematik (Modell 6f)	167
Abbildung 22: Zwei-Ebenen-Modell der kognitiven und non-kognitiven Faktoren und des SES	170
Abbildung 23: Parsimonischstes Zwei-Ebenen-Modell der kognitiven und non-kognitiven Faktoren, sowie des SES in Deutsch und Mathematik (Modell 7/ Modell 8a)	172

Tabellenverzeichnis

Tabelle 1: Übersicht der an der Studie beteiligten Schulen, der entsprechender Schülerzahl und der Rücklaufquote	90
Tabelle 2: Altersstruktur und Geschlechterverteilung der Stichprobe	91
Tabelle 3: Höchster Schulabschluss der Eltern in der Stichprobe im Vergleich zu Angaben des Datenreports 2011 (DESTATIS, 2011)....	92
Tabelle 4: Deskriptive Statistiken der kognitiven Fähigkeiten	121
Tabelle 5: Korrelative Betrachtung der Binnenstruktur der kognitiven Prädiktoren sowie der Interkorrelationen	122
Tabelle 6: Deskriptive Statistiken der non-kognitiven Faktoren	124
Tabelle 7: Konvergente und divergente Korrelationen der non- kognitiven Faktoren innerhalb einer Domäne und über die Domänen hinweg	125
Tabelle 8: Verteilungskennwerte und korrelative Zusammenhänge des höchsten elterlichen Bildungsabschlusses	126
Tabelle 9: Korrelative Zusammenhänge der kognitiven und non- kognitiven Prädiktoren und des SES	127
Tabelle 10: Deskriptive Statistiken der Schulnoten	129
Tabelle 11: Korrelationen der Schulnoten und des fachspezifischen Notendurchschnitts	131
Tabelle 12: Korrelative Zusammenhänge der Prädiktoren mit dem fachspezifischen Notendurchschnitt	132
Tabelle 13: Regression der mittleren Leistung in Deutsch auf WM und gf	134
Tabelle 14: Regression der mittleren Leistung in Mathematik auf WM und gf.....	136
Tabelle 15: Fit-Indizes der Prognosemodelle der kognitiven Prädiktoren und des SES in Deutsch und Mathematik	146
Tabelle 16: Fit-Indizes der Zwei-Ebenen-Modelle der kognitiven Prädiktoren und des SES in Deutsch und Mathematik	152
Tabelle 17: Standardisierte Pfadgewichte und latente Korrelationen der Modelle mit den Prädiktoren WM, gf und SES getrennt nach Schulfach.....	153
Abbildung 18: Parsimonischstes Prognosemodell der kognitiven und non-kognitiven Faktoren, sowie des SES in Deutsch (Modell 5f).....	163

Tabelle 19: Fit-Indizes der Prognosemodelle mit kognitiven und non-kognitiven Faktoren, sowie SES für Deutsch und Mathematik.....	168
Tabelle 20: Fit-Indizes der Zwei-Ebenen-Modelle mit kognitiven und non-kognitiven Faktoren, sowie SES in Deutsch und Mathematik...	171
Tabelle 21: Standardisierte Pfadgewichte und latente Korrelationen der Modelle mit kognitiven und non-kognitiven Faktoren, sowie SES getrennt nach Schulfach	174
Tabelle 22: Im Kinder- und Elternfragebogen erfasste Konstrukte mit jeweiliger Erhebungsperspektive.....	247
Tabelle 23: Gängige "Daumenregeln" zur Bewertung der Modellgüte (in Anlehnung an Schermelleh-Engel et al., 2003)	247
Tabelle 24: Notation der Variablen im Pfadmodell der SEM (in Anlehnung an Weiber & Mülhhaus, 2010)	248
Tabelle 25: Vergleich der Faktorladungen und Korrelationen des Messmodells non-kognitiver Faktoren mit und ohne korrelierte Fehlerterme	249
Tabelle 26: Faktorladungen des jeweils parsimonischsten Modells der kognitiven Faktoren und SES.....	249
Tabelle 27: Faktorladungen des jeweils parsimonischsten Modells der kognitiven und nonkognitiven Faktoren, sowie SES	250
Tabelle 28: Korrelationen der untersuchten Konstrukte auf Item- bzw. Subskalenniveau.....	251

Abkürzungsverzeichnis

AIC	Akaike's Information Criterion
BFLPE	Big Fish Little Pond Effect
CFI	Comparative Fit Index
CFT 20-R	Culture Fair Test 20-R
CMIN	χ^2 - Wert
D	Notendurchschnitt Deutsch
DESTATIS	Statistisches Bundesamt Deutschland
df	Freiheitsgrad
DLPFC	Dorsolateraler präfrontaler Cortex
EM	Estimation Maximization
FSK	Fähigkeitsselbstkonzept
FSKd	Fähigkeitsselbstkonzept Deutsch
FSKm	Fähigkeitsselbstkonzept Mathematik
<i>g</i>	<i>g</i> -Faktor der allgemeinen Intelligenz
gc	kristalline Intelligenz
GEW	Gewerkschaft Erziehung und Wissenschaft
gf	fluide Intelligenz
GFI	Goodness of Fit Index
HAWIK IV	Hamburg-Wechsler-Intelligenztest für Kinder IV
IM	Intrinsische Motivation
IMd	Intrinsische Motivation Deutsch
IMd	Intrinsische Motivation Mathematik
KZG	Kurzzeitgedächtnis
LOGIK	Münchner Longitudinalstudie zur Genese individueller Kompetenzen
M	Notendurchschnitt Mathematik
MAR	Missingness at random
MCAR	Missingness completely at random
ML	Maximum Likelihood
p	Signifikanzniveau
PFC	Präfrontaler Cortex
PISA	Programme for International Student Assessment
PL	Phonologische Schleife

RMSEA	Root Mean Square Error of Approximation
SDT	Self-determination theory
SEM	Strukturgleichungsanalyse (structural equation modeling)
SES	Sozioökonomischer Status
VSSP	Visuo-spatial sketchpad
WM	Arbeitsgedächtnis (working memory)
ZE	Zentrale Exekutive

Einleitung

Es wird in unserer Gesellschaft als selbstverständlich erachtet, dass Menschen miteinander um Ressourcen, Zensuren, Studien- oder Arbeitsplätze konkurrieren. Dabei gilt der Leitgedanke, dass die jeweils Fähigsten sich schließlich durchsetzen werden. Doch nicht nur das Berufsleben ist stark durch Leistungsdenken geprägt, wonach das Streben nach Erfolg als Voraussetzung für einen arrivierten Werdegang und damit als positive Eigenschaft bewertet wird. Auch im schulischen Umfeld zeigen Redewendungen wie „ohne Fleiß kein Preis“, dass allgemein die Auffassung vertreten wird, dass ein Jeder selbstverantwortlich nach höchstmöglicher Leistung streben sollte, um durch Einsatz und Engagement zum Erfolg zu gelangen.

Doch was macht Menschen leistungsfähig? Und ist Leistungsstreben tatsächlich Kern des Erfolgs? Übertragen auf den schulischen Kontext stellt sich die Frage: Welche Schüler erweisen sich in unserem Schulsystem als erfolgreich, und warum? Mit dieser Frage beschäftigt sich die vorliegende Studie.

Dass der Bildungsforschung allgemein ein hoher gesellschaftlicher Wert beigemessen wird, zeigt sich beispielsweise in der Etablierung des Rahmenprogramms zur Förderung der empirischen Bildungsforschung im Jahr 2007, finanziert durch das Bundesministerium für Bildung und Forschung. Die Relevanz einer fundierten Bildungsforschung wird mit der Notwendigkeit begründet, die Entwicklung des Bildungssystems voranzutreiben und die Stellung Deutschlands in Wissenschaft und Forschung zu stärken (Bundesministerium für Bildung und Forschung, 2011). Wie aktuell die Bemühungen um eine Weiterentwicklung des Bildungssystems sind, zeigen die Diskussionen um eine Verlängerung der Grundschulzeit oder die Abschaffung des „Sitzenbleibens“. Auch die im Saarland aktuell beschlossene Einführung der Gemeinschaftsschule als Ersatz für Haupt- und Realschule, veranschaulicht die Dynamik im Bildungssystem. Dabei verfolgen alle Debatten und Neuerungen das Ziel, eine Lernumwelt zu schaffen, in der Schüler effektiv und ihren Fähigkeiten entsprechend Lernpotenziale optimal entfalten können. Dazu bedarf es jedoch fundierter empirischer Untersuchungen, um die Parameter aufzudecken, deren Manipulation sich als dienlich im Sinne der Zielverwirklichung erweist. So kann die Suche nach den Bedingungsfaktoren schulischer Leistung als eine der Kernfragen bezeichnet werden. Auch die vorliegende Arbeit zielt darauf ab,

einen Beitrag zu dieser Fragestellung zu leisten, durch die Untersuchung der relativen Bedeutung kognitiver und non-kognitiver Individualfaktoren, sowie des elterlichen sozioökonomischen Status auf die Vorhersage des Schulerfolgs.

Lange Zeit galt die allgemeine Denkfähigkeit oder Intelligenz, als die limitierende Größe in Bezug auf schulische Leistungen. Je intelligenter ein Kind desto besser die Schulleistung. So konnte der Zusammenhang zwischen allgemeiner Intelligenz und Schulerfolg vielfach empirisch nachgewiesen werden (Gustafsson & Undheim, 1996). Entsprechende Korrelationen liegen im Mittel bei .50, womit sich Intelligenz als stärkster Prädiktor schulischer Leistungen etabliert hat (z.B. Neisser et al., 1996). Die Zusammenhänge variieren je nach Operationalisierung des Leistungskriteriums, da die Betrachtung von Testergebnissen in standardisierten Leistungstests üblicherweise zu höheren Werten führt im Vergleich zu Studien, in denen Zensuren als Kriterium berücksichtigt werden (Deary, Strand, Smith & Fernandez, 2007).

Als Beleg für die bedeutende Rolle kognitiver Fähigkeiten hinsichtlich schulischer Leistungen sind weiterhin die aktuellen Befunde zu nennen, die den Zusammenhang zwischen Arbeitsgedächtnis und Schulerfolg bzw. verschiedensten Kernkompetenzen schulischer Leistungen bestätigen (z.B. Alloway et al., 2005). So gilt Arbeitsgedächtnis als wichtige Voraussetzung für Lernprozesse, weshalb ein verstärktes Forschungsinteresse hinsichtlich dieses Konstrukts und seiner Korrelate festzustellen ist. Besonders im Kontext der Diagnostik von Lernschwierigkeiten und der Suche nach geeigneten Interventionen erweist sich die Arbeitsgedächtnisforschung als vielversprechend (Carretti, Borella, Cornoldi & De Beni, 2009). Außerdem werden die Assoziation von Intelligenz und Arbeitsgedächtnis bzw. die zugrunde liegenden Mechanismen stark beforscht. Aufgrund vielfältiger Belege der Relevanz des Arbeitsgedächtnisses im Schulkontext findet diese Variable auch in der vorliegenden Studie Berücksichtigung.

Doch trotz des starken Einflusses kognitiver Fähigkeiten, reicht eine Fokussierung auf diese nicht auf, um das komplexe Geschehen zu verstehen, auf welchem schulische Erfolge basieren. Das Phänomen der Under- bzw. Overachiever, die unter der zu erwartenden Leistung zurückbleiben bzw. diese übersteigen, verdeutlicht die Tatsache, dass nicht nur „Köpfchen“ darüber entscheidet, wie erfolgreich sich eine Schullaufbahn gestaltet. So postulieren moderne Schulerfolgsmodelle die multiple Determiniertheit des Schulerfolgs (z.B. Helmke & Schrader, 2006) und fordern die Berücksichtigung zusätzlicher Bedingungsfaktoren, die über die kognitiven Faktoren hinausgehen. Auch wenn nach Sauer (2006) ein zu testendes Vorhersagemodell

stets unvollständig bleibt, da nie alle Einflussfaktoren gleichzeitig erfasst werden können, ist dennoch aus pädagogischer Sicht eine möglichst breite Erfassung verschiedener Prädiktoren erstrebenswert. Schließlich wäre ein Schulsystem wünschenswert, welches es jedem Kind ermöglicht, seine individuelle Bestleistung optimal abzurufen, ohne das Zurückbleiben Einzelner aufgrund fehlerhafter Lehrmeinungen über Grundlagen des Schulerfolgs und daraus abgeleiteter unangemessener pädagogischer Praxis.

Im Hinblick auf wichtige non-kognitive Faktoren hat sich gezeigt, dass besonders motivationale Variablen einen Einfluss ausüben können. Der Wille, das vorhandene Leistungspotenzial auszuschöpfen, ist ebenso von Bedeutung wie die eigene Einschätzung der Fähigkeit in einem bestimmten Bereich (Helmke & Weinert, 1997). Unter Berücksichtigung bisheriger Erkenntnisse werden daher das fachspezifische Fähigkeitsselbstkonzept, sowie die intrinsische Motivation als Repräsentanten der non-kognitiven Variablen in der vorliegenden Studie berücksichtigt. Besonders der inkrementelle Vorhersagebeitrag steht dabei im Fokus, also die Frage nach dem Varianzanteil, den die non-kognitiven Faktoren über die kognitiven Fähigkeiten hinaus zusätzlich aufklären können. Diesbezügliche Befunde geben Einblicke hinsichtlich der relativen Bedeutung der beiden Prädiktorgruppen.

Gerade der wichtige Aspekt der inkrementellen Validität wurde in vielen empirischen Studien bislang nicht berücksichtigt. So wurden häufig verschiedene Prädiktoren lediglich separaten Analysen unterzogen. Doch um zu einem Verständnis der real wirksamen Zusammenhänge zu gelangen, ist die Bewertung der relativen Bedeutung verschiedener Faktoren unabdingbar. Besonders die in der vorliegenden Studie umgesetzte gleichzeitige Betrachtung von Arbeitsgedächtnis, Intelligenz und Motivation stellt dabei einen neuen Ansatz dar, durch den bisherige Forschungsbefunde ergänzt werden können.

Selbstverständlich beeinflussen nicht nur individuelle Faktoren die schulischen Erfolge eines Kindes. Auch die Lernumwelt, die ein Kind umgibt, beispielsweise das Elternhaus, spielt eine Rolle. In diesem Zusammenhang herrscht eine rege Diskussion bezüglich der Chancengleichheit im Bildungssystem für Kinder aus unterschiedlichen sozialen Schichten. Die Aktualität dieser Debatte zeigt die Veröffentlichung eines Artikels der Gewerkschaft Erziehung und Wissenschaft (GEW) mit dem Titel „Arbeiterkinder müssen Superschüler sein ... um nach der Grundschule aufs Gymnasium zu gehen“ (Demmer, 2009). Darin wird der weit verbreiteten Sorge Ausdruck verliehen, dass bei der Vergabe von Empfehlungen für weiterführende

Schulen Kinder nicht gemäß ihren Fähigkeiten bewertet werden könnten, sondern die soziale Stellung der Eltern zu stark das Urteil der Lehrer beeinflusst. Zwar ist empirisch erwiesen, dass ein Zusammenhang zwischen sozioökonomischem Status (SES) und Intelligenz besteht (Neisser et al., 1996), so dass Zusammenhänge zwischen SES und Schulerfolg immer vor dem Hintergrund der Vererbung der elterlichen Intelligenz an die Kinder zu diskutieren sind. Dennoch wird der SES aufgrund der gesellschaftlichen Relevanz als Prädiktor in die vorliegende Studie einbezogen. Schließlich gibt es auch Hinweise auf eine inkrementelle Varianzaufklärung des SES über Intelligenz hinaus (Schick et al., 2006).

Für diese Art der Fragestellung eignet sich besonders die Methode der Strukturgleichungsanalyse, welche es ermöglicht, verschiedene Prädiktoren gleichzeitig hinsichtlich ihrer Vorhersagekraft zu untersuchen und zu vergleichen. Eine fachspezifische Operationalisierung erlaubt dabei die Untersuchung differentieller Effekte der Fächer Deutsch und Mathematik. Beide Aspekte finden in der vorliegenden Studie Berücksichtigung.

Die Untersuchung ist eingebettet in das internationale Graduiertenkolleg IRTG *Adaptive Minds* 1457, welches von der Deutschen Forschungsgesellschaft (DFG) gefördert wird. Im Rahmen der unter Projekt F geführten Studie, erfolgte eine Datenerhebung an zehn verschiedenen saarländischen Grundschulen, wobei insgesamt 320 Kinder der vierten Klassenstufe teilnahmen. Die kognitive Testung fand als Gruppentestung in den jeweiligen Schulen statt. Mittels Fragebogen wurden die non-kognitiven Faktoren sowie der elterliche sozioökonomische Status erfasst. Parallel zur Erhebung in Deutschland fand die Testung einer chinesischen Stichprobe in Peking statt, um Unterschiede hinsichtlich der relativen Bedeutung der verschiedenen Prädiktoren zwischen den Kulturen herausstellen zu können. Da dieser kulturvergleichende Aspekt des Projekts F jedoch nicht Gegenstand der vorliegenden Dissertation ist, wird auf die chinesische Datenerhebung nicht näher eingegangen.

Die Entscheidung, eine Stichprobe von Viertklässlern zu untersuchen, basiert vorrangig auf der Tatsache, dass die vierte Klasse eine bedeutende Schnittstelle im deutschen Schulsystem darstellt. So wird in dieser Klassenstufe darüber entschieden, welche weiterführende Schule nach der Grundschulzeit besucht werden soll. Der Wechsel von der Grund- zur weiterführenden Schule stellt eine Herausforderung für Kinder und Eltern dar. Schließlich ändern sich zahlreiche

Parameter der Lernumgebung, sowie die Zusammensetzung der Klasse als Referenzgruppe, was Auswirkungen auf motivationale Aspekte wie Fähigkeitsüberzeugungen hat (z.B. Marsh, 1987a). Auch die stärkere Belastung durch steigende Stundenzahlen, neue Schulfächer und vermehrte Hausaufgaben erfordert ein hohes Maß an Anpassungsfähigkeit seitens der Kinder. Aufgrund der weitreichenden Bedeutung, die die Wahl der weiterführenden Schule für den künftigen Bildungsweg und spätere Karrierechancen hat, ist es wichtig, die Erfolgsfaktoren zu untersuchen, die zwischen Kindern mit guten und schlechten Lernleistungen zum Ende der Grundschulzeit differenzieren. Da die Entscheidung, welche weiterführende Schule besucht werden soll, vorrangig auf den Zensuren der vierten Klasse basiert, werden Schulnoten als Erfolgskriterium herangezogen. Damit wird ein Kriterium operationalisiert, welches sehr nah am realen Schulalltag der Kinder ist und damit unmittelbare Relevanz besitzt (Steinmayr & Spinath, 2009).

Der erste Teil dieser Arbeit fasst die theoretischen Grundlagen der Studie zusammen. So erfolgt zunächst die Beschreibung wichtiger Aspekte, die das Konzept Arbeitsgedächtnis betreffen. Im Anschluss daran wird eine zusammenfassende Beschreibung bezüglich Definition und Struktur von Intelligenz gegeben, bevor die Darstellung wichtiger Aspekte der Motivationsforschung erfolgt. Die Definition und Bedeutung des sozioökonomischen Status folgt anschließend. Der theoretische Überblick schließt mit Ausführungen bezüglich des Kriteriums Schulerfolg.

Im zweiten Teil der Arbeit erfolgt die Beschreibung der Fragestellung unter Berücksichtigung des dargelegten theoretischen Hintergrunds und der empirischen Befunde, sowie die Ableitung der Hypothesen dieser Studie.

Der dritte Teil stellt die Studienergebnisse dar. Nach der Veranschaulichung der deskriptiven Ergebnisse und korrelativen Zusammenhänge, werden Befunde der Regressions- und Kommunalitätenanalysen berichtet, welche die kognitiven Faktoren Arbeitsgedächtnis und Intelligenz berücksichtigen. Im Anschluss daran erfolgt die Darstellung der Ergebnisse der Strukturgleichungsmodellierungen, sowohl für die Modelle, welche singular kognitive Prädiktoren einschließen, als auch für solche, die gleichzeitig kognitive und non-kognitive Faktoren berücksichtigen.

Im Anschluss an die Ergebnisdarstellung werden die Befunde diskutiert. Hierbei wird zunächst auf spezifische Befunde auf Konstruktebene eingegangen. Im Anschluss werden die Befunde bezüglich des Vergleichs der prädiktiven Validität von

Arbeitsgedächtnis und Intelligenz diskutiert. Daran anschließend wird die relative Bedeutung der kognitiven und non-kognitiven Prädiktoren im Hinblick auf die Vorhersage des Schulerfolgs diskutiert, bevor schließlich auf die Vorhersagekraft des SES eingegangen wird.

Anschließend werden die Ergebnisse resümiert, bevor die Arbeit mit einer Darstellung der Einschränkungen und einem Ausblick bzw. dem Hinweis auf praktische Implikationen endet.

I Theoretische und empirische Grundlagen

Im folgenden Abschnitt erfolgt die Darstellung des theoretischen Hintergrunds der in dieser Arbeit betrachteten Determinanten des Schulerfolgs. So beschäftigt sich Kapitel 1.1 mit dem Konstrukt Arbeitsgedächtnis. In Kapitel 1.2 werden Modelle und Befunde bezüglich der allgemeinen kognitiven Fähigkeit vorgestellt. Im Anschluss werden wichtige Konzepte und Modellvorstellungen der Motivationsforschung beschrieben (Kapitel 1.3). Dabei liegt der Schwerpunkt auf den zentralen Faktoren Fähigkeitsselbstkonzept und intrinsische Motivation, welche in der vorliegenden Arbeit als Repräsentanten non-kognitiver Einflussfaktoren fungieren. Schließlich erfolgt eine kurze Darstellung wichtiger Aspekte des sozioökonomischen Status als Einflussfaktor auf schulische Leistungen (Kapitel 1.4), bevor der theoretische Überblick mit einer Beschreibung des Konzepts Schulerfolg, sowie diesbezüglich relevanter Befunde schließt (Kapitel 1.5).

1.1 Arbeitsgedächtnis

Im folgenden Kapitel werden die grundlegenden theoretischen Aspekte bezüglich des Konstrukts *Arbeitsgedächtnis* (working memory, WM) dargelegt. Nach einer kurzen Einführung (Kapitel 1.1.1) erfolgt die Darstellung dreier wichtiger Modellvorstellungen in Abschnitt 1.1.2. Es handelt sich hierbei um das Arbeitsgedächtnismodell von Baddeley (1974, 1986, 2000), auf welches sich diese Arbeit stützt, sowie die ebenfalls einflussreichen Modelle von Cowan (1988) bzw. von Engle, Kane und Tuholski (1999). Kapitel 1.1.3. befasst sich mit der Organisation des Arbeitsgedächtnissystems. Anschließend werden in Kapitel 1.1.4 entwicklungsbedingte Veränderungen des Arbeitsgedächtnisses im Kindesalter dargestellt. Es wird mit einer Zusammenfassung der wichtigsten Kernaspekte geschlossen (Kapitel 1.1.5).

1.1.1 Einführung

Aus der Sichtung der bereits existierenden Literatur bezüglich des Themas Arbeitsgedächtnis wird schnell ersichtlich, dass es sich hierbei um ein sehr umfangreiches, komplexes Forschungsgebiet handelt. Es existieren verschiedene Ansätze mit teils konträren Modellvorstellungen, so dass eine übergreifende, strikte Definition des Konzeptes bislang fehlt. Allgemein bezeichnet der Begriff jene kognitiven Prozesse und Strukturen, die uns befähigen, die zur Bearbeitung einer Aufgabe notwendigen Informationen zeitweilig aufrecht zu erhalten und zu manipulieren (Alloway, 2009). Dieser Definition folgend spielt WM als Fundament kognitiver Fähigkeiten eine bedeutsame Rolle für alle höheren kognitiven Leistungen (Rummer & Fiebach, 2010). So wird angenommen, dass unterschiedliche Arbeitsgedächtnisleistungen auch mit Leistungsunterschieden in anderen kognitiven Fähigkeiten, wie etwa der fluiden Intelligenz einhergehen (Jarrold & Towse, 2006). Neben den Zusammenhängen zu bestimmten kognitiven Fähigkeiten konnten außerdem Verbindungen zu bestimmten Rechen- und Leseschwierigkeiten (Bull & Scerif, 2001; Swanson & Berninger, 1995), bzw. zu allgemeinen Lernschwierigkeiten (Siegel & Ryan, 1989; Swanson, 1993) nachgewiesen werden. Das immense Interesse an dem Konzept WM erklärt sich zum Teil auch durch die Bedeutung dieser kognitiven Fähigkeit im alltäglichen Leben. Schließlich erfordern auch Handlungen wie Auto fahren oder Kopfrechnen das Speichern, Manipulieren und Integrieren von Information.

Ursprünglich aus der kognitiven Psychologie stammend, findet das Konstrukt WM zwischenzeitlich Anwendung in den unterschiedlichsten Teilgebieten der Psychologie, sei es klinische, soziale, Entwicklungs-, oder pädagogische Psychologie. Jeder dieser Forschungszweige profitierte von der Einbindung dieses Konzepts (Conway et al., 2005). In der klinischen Forschung wurde beispielsweise eine Beziehung zwischen WM und Depression nachgewiesen (Arnett et al., 1999), in der Neuropsychologie wird Arbeitsgedächtniskapazität als Marker für früh einsetzende Alzheimersche Krankheit diskutiert und die Entwicklungspsychologie betont die Bedeutung der Entwicklung des WMs für die Entwicklung des Spracherwerbs (Alloway et al., 2005). Weiterhin mehren sich die Befunde, dass WM als Prädiktor schulischer Leistung eine ähnlich bedeutsame Rolle spielt, wie die in der Bildungsforschung etablierte Variable der allgemeinen Intelligenz (Luo, Thompson & Detterman, 2006), weshalb diesem Konstrukt in der vorliegenden Arbeit ein hoher Stellenwert beigemessen wird. Zwar zählt die Arbeitsgedächtnisforschung

nicht zu den klassischen Schwerpunkten der differentiellen Psychologie, aber dennoch hat diese die theoretische Entwicklung des Konzepts deutlich beeinflusst. Nicht nur hinsichtlich der Frage inwiefern WM als Korrelat anderer höherer kognitiver Fähigkeiten verstanden werden kann und welche Prozesse dieser Beziehung zu Grunde liegen, sondern auch bezüglich der Klärung der inneren Organisation des Arbeitsgedächtnissystems hat sich die Betrachtung interindividueller Differenzen bewährt (Jarrold & Towse, 2006).

Wie zu Beginn bereits erwähnt, existieren unterschiedliche Konzeptionen im Hinblick auf Struktur und Organisation des WM-Systems (Miyake & Shah, 1999). Im Rahmen dieser Arbeit können nicht alle Ansätze erschöpfend dargestellt werden, so dass lediglich drei einflussreiche Konzeptionen herausgegriffen und im Folgenden erläutert werden. Es handelt sich dabei um das Multi-Komponenten-Modell von Baddeley (Baddeley & Hitch, 1974), das Embedded-Process-Model nach Cowan (1999) und um das Modell von Engle, Kane und Tuholsky (1999). Die vorliegende Arbeit stützt sich besonders auf das Modell von Baddeley, da dieses gut erforscht und in der Lernforschung besonders etabliert ist (Berti, 2010).

1.1.2 Modellvorstellungen

Allgemein können die verschiedenen Arbeitsgedächtnismodelle danach klassifiziert werden, inwiefern sie WM als domänen-spezifische oder –unspezifische Instanz betrachten. Shah und Miyake (1996) verfolgen beispielsweise einen domänen-spezifischen Ansatz, indem sie unterscheidbare Arbeitsgedächtniskomponenten postulieren, die entweder für verbalen Input oder für visuell-räumliche Informationen verantwortlich sind. Modelle, die sowohl domänen-spezifische als auch –unspezifische Anteile annehmen (Baddeley, 2000; Engle, Kane & Tuholsky, 1999) postulieren hingegen, dass das WM-System eine domänen-übergreifende Instanz beinhaltet, welche allerdings die Arbeit von domänen-spezifischen Subkomponenten überwacht und steuert. Im Folgenden sollen die drei einflussreichsten Modelle dieser Art näher beschrieben werden.

1.1.2.1 Das Multi-Komponenten-Modell von Baddeley (Baddeley & Hitch, 1974)

Das Baddeley'sche Modell kann als eines der ältesten und gleichzeitig einflussreichsten Modelle der Arbeitsgedächtnisforschung bezeichnet werden. Bis heute erfreut es sich aufgrund seiner relativen Einfachheit einer großen Beliebtheit in der Grundlagen- und angewandten Forschung. In seinen frühen Forschungen beschäftigte sich Baddeley zunächst mit dem *Kurzzeitgedächtnis* (KZG), auch „immediate memory“ (Ackerman, Beier & Boyle, 2005, S. 33) genannt, worunter ein passives Speichersystem verstanden wurde. Innovativ an seinem Vorgehen war nun, dass er erstmals Aufgaben einsetzte, die sich vom bis dahin üblichen Gedächtnisspannen-Paradigma dahingehend unterschieden, dass zwei Aufgaben gleichzeitig auszuführen waren, beispielsweise eine einfache Spannaufgabe verknüpft mit einer Entscheidungsaufgabe (Baddeley, 1998). Aus den Befunden schlussfolgerte Baddeley, beeinflusst durch Arbeiten zum multimodalen Modell von Atkinson und Shiffrin (1968), dass zur Bearbeitung dieser Art von Aufgaben statt eines passiven Kurzzeitgedächtnisspeichers (Baddeley, 2006) ein aktives Gedächtnissystem genutzt werden müsse, bestehend aus einzelnen interagierenden Teilkomponenten. Diesen aktiven Speicher bezeichneten Baddeley und Hitch (1974) als *working memory*¹ und verstanden darunter die Basis aller kognitiver Funktionen. Die Arbeiten von Baddeley und Hitch (1974) gelten als Meilenstein in der Gedächtnisforschung, da ihre funktionale Betrachtung des Gedächtnisses zu einem besseren Verständnis der Prozesse und Charakteristika des Gedächtnissystems beitrug (Jarrold & Towse, 2006).

Im Mittelpunkt des *Multi-Komponenten-Modells* steht die *zentrale Exekutive* (ZE), eine flexible Kontrollinstanz, die die Verantwortung für eine Vielzahl regulatorischer Aufgaben trägt, wie etwa Aufmerksamkeits- und Handlungskontrolle, Kontrolle des Informationsflusses oder Regulierung des Zugangs und Abrufs von Informationen aus anderen Gedächtnissystemen, beispielsweise dem Langzeitgedächtnis (Baddeley, 1986; Baddeley, Emslie, Kolodny, & Duncan, 1998). Um ihre vielfältigen Aufgaben zu erfüllen steht der ZE allerdings nur ein begrenzter Umfang an flexibel einsetzbarer Kapazität zur Verfügung (Baddeley & Hitch, 1974). Neuerdings wird vermehrt diskutiert, wie bedeutsam Aufmerksamkeitsprozesse für das Funktionieren

¹ Der Begriff *working memory* wurde bereits von Miller, Galanter und Pribram (1960) verwendet.

der ZE sind, wie die Modelle von Cowan (1988) und Engle et al. (1999) in den folgenden Abschnitten verdeutlichen werden. Auch im Baddeley'schen Modell wird die Bedeutung der ZE für Aufmerksamkeitskontrollprozesse formuliert. Kritisch hierbei ist jedoch, dass es sich keineswegs um einheitliche, klar umschriebene Prozesse handelt, wodurch die Klärung der Beziehung erschwert wird. Verschiedene Aspekte der aktuellen Aufmerksamkeitsforschung (z.B. geteilte Aufmerksamkeit, Aufmerksamkeitswechsel) können zwar durch die Annahme einer ZE besser beleuchtet werden, dennoch bleibt festzuhalten, dass es weiterer Forschung bedarf, um die Beziehung zu konkretisieren (Baddeley & Logie, 1999). Offen bleibt außerdem, wie die Regulation und Organisation der ZE tatsächlich vonstattengeht. So stellt sich die Frage, ob allen Funktionen der ZE eine gleich große Bedeutung zukommt, oder ob eine Hierarchie innerhalb des Kontrollsystems existiert. Baddeley und Logie (1999) lassen diese Frage unbeantwortet: „we leave open to investigate the question of whether the central executive resembles an organization run by a single chairperson or one governed by the collective wisdom of a committee of equals“ (S.40). Die wichtigste Aufgabe der ZE besteht jedoch darin, die Arbeit der untergeordneten Slavesysteme zu koordinieren und zu kontrollieren, welche im Anschluss näher spezifiziert werden sollen.

Die Slave- oder Subsysteme stehen in direkter Verbindung zur ZE und stellen temporäre Gedächtnisspeicher da. Es wird zwischen der *phonologischen Schleife* (PS), die auf die Verarbeitung sprachbasierter verbaler Information spezialisiert ist, und dem *visuo-spatial sketchpad* (VSSP) differenziert, welcher visuell-räumlichen Input verarbeitet. Die PS unterteilt sich weiter in einen Kurzzeitspeicher und einen Rehearsalmechanismus. Der Speicher dient der Aufrechterhaltung verbaler Repräsentationen. Hierzu werden verbale Charakteristika genutzt, allerdings verfällt der im Speicher gehaltene Input sehr schnell, nämlich nach ca. zwei Sekunden. Ein verlängertes Halten der Information kann mit Hilfe des subvokalen Wiederholens (Rehearsal) erzielt werden, einem seriellen Mechanismus, der den Speicher unterstützt. Bezüglich des VSSP wird eine Aufteilung in eine *passiv-visuelle* (inner eye) und *aktiv-räumliche* (inner scribe) Komponente diskutiert (Della Sala, Gray, Baddeley, Allamano & Wilson, 1999; Farah, Hammond, Levine, & Calvanio, 1988; Hanley, Young, & Pearson, 1991; Logie, 1995; Logie & Pearson, 1997). Die domänen-spezifischen Subsysteme interagieren sowohl untereinander als auch, wie weiter oben erläutert, mit der ZE. Im Jahr 2000 erweiterte Baddeley seine Konzeption um eine weitere Komponente, den *episodic buffer*. Diese ebenfalls kapazitätsbegrenzte Komponente nutzt multidimensionale Codes, die aus

verschiedenen Komponenten des WMs stammen können, um sie mit Repräsentationen des Langzeitgedächtnisses zu verknüpfen. Der episodic buffer bildet also durch Verknüpfen von Input aus verschiedenen Domänen und Gedächtnissubsystemen Informationsketten (Alloway, Gathercole & Pickering, 2006). Alloway, Gathercole, Willis und Adams (2004) weisen darauf hin, dass diese episodischen Verknüpfungen möglicherweise die Basis unseres bewussten Erlebens darstellen könnten.

Allgemein ermöglicht WM es Informationen für eine gewisse Zeit im Gedächtnis zu halten und gleichzeitig zu transformieren (Alloway, 2009). Wichtig hierbei ist, dass nur die Subsysteme eine Speicherfunktion besitzen, nicht jedoch die ZE, welche die Arbeit der Subsysteme überwacht. Allerdings kann die ZE die Speicherfähigkeit über die Kapazität der Subsysteme hinaus erweitern, indem sie Langzeitgedächtnisinhalte aktiviert oder auf andere Subsysteme zugreift. Dennoch ist die Kapazität des WMs limitiert. Information verfällt, wenn die Menge an Items, die es zu verarbeiten gilt die individuelle Kapazitätsgrenze überschreitet oder die Zeitspanne, die die Information gehalten werden muss, zu lange andauert (Ackerman, Beier & Boyle, 2002).

Die Fähigkeit der simultanen Speicherung und Transformation von Information ist von solch weitreichender Relevanz, dass WM mit allen komplexen kognitiven Aufgaben in Zusammenhang steht, die auf der Koordination und Speicherung von Information beruhen, wie z.B. Reasoning oder Sprachverständnis (Ackerman et al., 2002). In Anbetracht der Vielzahl an Aufgaben, für welche Schüler gleichzeitig Information speichern und transformieren müssen, liegt es auf der Hand, dass WM auch im Schulkontext von großer Bedeutung sein muss, sei es beim Lesen, Kopfrechnen oder Schreiben von Aufsätzen. Hierauf wird in Kapitel 1.5.1 näher eingegangen.

Bezüglich der Messung der Arbeitsgedächtniskapazität postulierte Baddeley (1986), dass Aufgaben, die simultanes Speichern und Transformieren erfordern, besonders geeignet seien. Dieser Grundsatz führte zur Entwicklung zahlreicher sogenannter *Storage-and-Processing-Tasks* (SAPTs), die dieser Anforderung genügen. Als Prototyp gilt die Reading Span- Aufgabe von Daneman und Carpenter (1980), die zur Vorhersage des Leseverstehens entwickelt wurde. Auch in der vorliegenden Arbeit wurden Aufgaben eingesetzt, die dem Paradigma der simultanen Speicherung und Transformation von Information folgen (siehe Kapitel 3.4.1).

Kritiker der Baddeley'schen Modellvorstellung bemängeln die stetig anwachsende Zahl der Funktionen, die mit der ZE in Verbindung gebracht werden. „The promiscuous way in which (the) central executive has acquired functions is potentially a substantial problem“ (Towse & Cowan, 2005, S. 14, zitiert nach Berch, 2008). Außerdem ist das Verhältnis von Studien mit verbalem und non-verbalem Stimulusmaterial unausgeglichen, da Untersuchungen mit sprachlichem Material dominieren (Berti, 2010). Nichtsdestotrotz wird das Modell durch zahlreiche Befunde gestützt. Diese stammen sowohl aus experimentellen Studien (z.B. Baddeley, Emslie, Kolodny & Duncan, 1998), als auch aus dem Bereich der Entwicklungspsychologie (z.B. Towse, Hitch & Hutton, 1998) oder der Neuropsychologie (z.B. Della Salla et al., 1999)

1.1.2.2 Das Embedded-Process-Model von Cowan (1999)

Das *Embedded-Process-Model* von Cowan (1999) stellt ein einflussreiches, prozessorientiertes Arbeitsgedächtnismodell dar, das insbesondere durch Befunde zum Zusammenhang von Aufmerksamkeit und Gedächtnis beeinflusst wurde. Im Gegensatz zu Baddeley beschäftigte sich Cowan überwiegend mit auditiver statt mit verbaler Verarbeitung. Das Modell beruht auf der Annahme eines hierarchisch organisierten Systems, bestehend aus Langzeitgedächtnis, aktiviertem Gedächtnis und Aufmerksamkeitsfokus. Das aktivierte Gedächtnis stellt eine zeitlich begrenzte Aktivierung einer Untermenge des Langzeitgedächtnisses dar, während der Aufmerksamkeitsfokus wiederum eine Untermenge dieser Teilmenge ist.

Die Kapazität des Aufmerksamkeitsfokus ist begrenzt und kann entweder willkürlich durch eine zentrale Kontrollinstanz oder unwillkürlich durch das Aufmerksamkeits-Orientierungs-System gesteuert werden. Aufmerksamkeit wirkt sich auf die Verarbeitung von Input dahingehend aus, dass sie bestimmt, welche und wie viele der wahrgenommenen Reize kodiert werden. Außerdem beeinflusst sie die Abrufbarkeit von Repräsentationen. Liegt Information im Fokus der Aufmerksamkeit, ist sie „qua Definition im Arbeitsgedächtnis“ (Berti, 2010, S.5). WM ist allerdings nicht auf die im Fokus der Aufmerksamkeit liegenden Inhalte begrenzt. Denn in Cowan's Modell wird WM funktional definiert und meint alle kognitiven Prozesse, durch welche Informationen in einem Zustand der Zugänglichkeit gehalten werden, um sie bei der Lösung von Aufgaben einzusetzen, wie z. B. beim Sprachverstehen, Problemlösen

oder bei Entscheidungsfindungsprozessen (Cowan, 1999). Es ist nicht möglich, solche komplexen Aufgaben lediglich mit Hilfe der wenigen im Aufmerksamkeitsfokus befindlichen Informationen zu lösen. Eine Unterstützung durch weitere aktivierte Quellen außerhalb der fokussierten Aufmerksamkeit und durch das Langzeitgedächtnis ist notwendig (Cowan, 1999). Cowan definiert WM also weniger als eigenständige Struktur, sondern vielmehr als funktionalen Zustand, der besonders durch Aufmerksamkeitsprozesse beeinflusst wird (Berti, 2010).

Die liberale Definition von WM nach Cowan wird jedoch von Kritikern als zu unspezifisch bemängelt. Denn überspitzt könnten selbst Papier und Bleistift als zum WM gehörend betrachtet werden, da auch Aufschreiben dem Ziel dient, Informationen im Zustand der Zugänglichkeit zu halten (Berti, 2010). Dennoch integriert das Modell wichtige Annahmen bezüglich der Assoziation von WM und Aufmerksamkeit und erweist sich im Hinblick auf die weitere Untersuchung dieser Beziehung als nützlich.

1.1.2.3 Das Modell von Engle, Kane und Tuholski (1999)

Die von Engle und Kollegen (1999) entwickelten Konzeptionen gehen auf die Frage zurück, wie und weshalb Intelligenz und WM zusammenhängen. Ähnlich wie Cowan's Modell beruht auch dieses Modell auf der Annahme, dass WM drei Teilbereiche umfasst: die über einen Schwellenwert aktivierten Langzeitgedächtnisinhalte, Prozesse und Fähigkeiten, die eine solche Aktivierung und deren Aufrechterhaltung ermöglichen, sowie kontrollierende Aufmerksamkeit als dritte grundlegende Komponente. Damit gemeint ist „the capacity for controlled, sustained attention in the face of interference or distraction” (Engle, Kane & Tuholsky, 1999, S. 104). Kontrollierende Aufmerksamkeit wird als domänen-unspezifisch und kapazitätsbegrenzt charakterisiert und dient entweder der Herstellung von Aktivierung (controlled retrieval), der Aufrechterhaltung von Aktivierung oder der Unterdrückung (inhibition) irrelevanter Reize. Damit ist diese Modellkomponente mit Baddeley's ZE vergleichbar. Neben dieser domänen-unspezifischen Instanz schließt auch das Modell von Engle et al. (1999) domänen-spezifische Codes und Rehearsalmechanismen ein, vergleichbar mit Baddeley's Subsystemen PL und VSSP. Allerdings ist nach Engle et al. (1999) Baddeley's Modell diesbezüglich nicht erschöpfend, denn die Autoren gehen davon aus, dass neben verbalen und visuell-

räumlichen Codes eine Vielzahl weiterer Codes möglich wäre. Laut Engle et al. (1999) stützen sowohl behaviorale, als auch neuropsychologische und neuroanatomische Befunde die Idee eines hierarchisch organisierten Arbeitsgedächtnis- bzw. Aufmerksamkeitssystems, welches einen domänen-unspezifischen Generalfaktor (Prozesskomponente) und mehrere domänen-spezifische Subsysteme (Speicherkomponente) vereint. Dieser Gedanke findet sich auch im Baddeley'schen Modell wieder. Neuroanatomisch scheint kontrollierende Aufmerksamkeit besonders mit dem *dorsolateralen präfrontalen Cortex* (DLPFC) in Verbindung zu stehen (Engle et al., 1999).

Das Verhältnis von WM und KZG spezifizieren Engle und Kollegen derart, dass KZG als Untermenge des WMs betrachtet wird. Beim KZG handelt es sich um einen einfachen Speicher, der die über einen Schwellenwert aktivierten Langzeitgedächtnisinhalte umfasst. WM hingegen ist „Kurzzeitgedächtnis plus kontrollierende Aufmerksamkeit“ (Berti, 2010, S.5), denn zum Halten der aktivierten Repräsentationen ist der Einsatz von Aufmerksamkeit erforderlich.

Es sei an dieser Stelle nochmals betont, dass Engle und Kollegen ihre Modellvorstellungen gezielt nutzen, um den Zusammenhang zwischen Intelligenz und WM zu beleuchten. Auch wenn interindividuelle Unterschiede sowohl in domänen-spezifischen, als auch in unspezifischen Arbeitsgedächtnisprozessen zu finden sind, sind laut Engle et al. (1999) besonders die Unterschiede in der Fähigkeit zur kontrollierten Verarbeitung für die Beziehung zur fluiden Intelligenz ausschlaggebend. So weisen Menschen mit unterschiedlicher psychometrischer Intelligenz auch Unterschiede in ihrer Fähigkeit zur kontrollierten Verarbeitung auf (Engle et al., 1999). „Controlled attention capabilities“ (Engle et al., 1999, S.112) sind notwendig, um Ziele aufrechtzuerhalten und um diese selbst angesichts störender Einflüsse und Distraktoren weiterzuverfolgen (Engle et al., 1999). Die Beziehung zwischen WM und Intelligenz wird ausführlicher in Kapitel 1.2.6 thematisiert.

1.1.3 Empirische Befunde zur Frage der Domänen-Generalität bzw. -Spezifität des Arbeitsgedächtnisses

Wie in Kapitel 1.1.1 angedeutet, stellt die Untersuchung der Organisation des Arbeitsgedächtnissystems eine Kernfrage des differentiellen Ansatzes in der Arbeitsgedächtnisforschung dar. Es wird dabei mit Hilfe korrelativer Methoden, sowie der Schätzung latenter Variablen versucht, Erkenntnisse über die Binnenstruktur des WM-Systems zu gewinnen. Besonders die Frage nach der Domänenspezifität bzw. -generalität wurde vielfach untersucht. Zur Klärung dieser Fragestellung dienen einerseits Studien, die die Zusammenhänge zwischen verschiedenen Arbeitsgedächtnisaufgaben untersuchen und andererseits solche, die die Leistung in Arbeitsgedächtnisaufgaben mit anderen kognitiven Leistungen in Verbindung setzen (Vock, 2005). Hier sprechen z.B. die Befunde von Shah und Miyake (1996) für die Domänenspezifitäts-Annahme, da in ihrer Studie die Performanz im verbalen *Scholastic Aptitude Test* (SAT) nur durch die Reading Span-Aufgabe vorhergesagt werden konnte, nicht jedoch durch visuell-räumliche Aufgaben. Für die visuell-räumliche Domäne fanden sich genau umgekehrte Zusammenhänge. In einem nächsten Schritt entwickelten Shah und Miyake (1996) sogenannte *domain-mismatching-tasks*, welche sich dadurch auszeichnen, dass durch Speicherung und Verarbeitung unterschiedliche Domänen angesprochen werden (beispielsweise Speichern verbalen Inputs bei Verarbeitung von visuell-räumlicher Information). Die Korrelation der Aufgabe mit anderen Tests zeigte, dass sich die Speicherdomäne als entscheidender für die Zusammenhänge erwies, als die Verarbeitungsdomäne, weshalb die Autoren schlussfolgerten, dass das Verarbeiten von Information eher domänen-unspezifischer Natur sei, während das Abspeichern einen domänen-spezifischen Vorgang darstellt.

Umgekehrt sprechen jedoch auch einige Studien für die Domänen-Generalitätshypothese. So konnte belegt werden, dass verbale Arbeitsgedächtnisaufgaben visuell-räumliche Fähigkeiten ebenso gut vorhersagen können wie verbale Fähigkeiten und umgekehrt (Bayliss, Jarrold, Gunn & Baddeley, 2003; Süß, Oberauer, Wittmann, Wilhelm & Schulze, 2002). Eine diesbezüglich wichtige Studie stammt von Swanson (1996), der eine Untersuchung an Kindern zwischen fünf und neunzehn Jahren durchführte und zeigen konnte, dass sich der Zusammenhang zwischen Schulleistung, Testleistung und WM als unabhängig vom Aufgabentyp erwies.

Untersuchungen, die den Zusammenhang verschiedener Arbeitsgedächtnisaufgaben betrachten, finden üblicherweise moderate bis hohe Korrelationen. Eine exploratorische Faktorenanalyse von Colom, Flores-Mendoza und Rebollo (2003) lieferte einen gemeinsamen Arbeitsgedächtnisfaktor für verschiedene Arbeitsgedächtnisaufgaben. Mittels konfirmatorischer Faktorenanalyse konnten Conway, Cowan, Bunting, Theriault und Minkoff (2002) zeigen, dass verschiedene komplexe numerische und verbale Spannaufgaben auf einen latenten Faktor luden, ebenso verschiedene einfache Spannaufgaben. Dies sprach laut den Autoren für eine Abgrenzung von Arbeits- und Kurzzeitgedächtnis, aber gegen eine innere Differenzierung nach Inhalten. Somit stützte diese Studie die Domänen-Generalitäts-Hypothese. Manko dieser Studie war jedoch der Mangel an visuell-räumlichen Aufgaben. Die Studie von Kane et al. (2004) umfasste sowohl verbale als auch räumliche Aufgaben. Hier fittete ein Ein-Faktor-Modell, in dem verbal-numerische und räumliche Aufgaben auf eine gemeinsame latente Variable luden, nicht signifikant schlechter als ein Zwei-Faktor-Modell mit getrennten verbal-numerischen und räumlichen Variablen. Auch die Befunde von Ackerman et al. (2002) sprechen für die Ein-Faktor-Hypothese, ebenso wie die Studien von De Jong und Das-Smaal (1995) und De Jong und de Jong (1996), in denen Kinder untersucht wurden. Süß (2001) fand in einer Studie Hinweise auf ein „hierarchisches, bimodales Strukturmodell des Arbeitsgedächtnisses“ (S.116). Strukturgleichungsmodelle ergaben einen Generalfaktor, sowie zwei getrennte Faktoren für figurale Inhalte bzw. für verbal-numerische. Dieser Befund stützt die Modellvorstellung von Baddeley.

Die Zusammenschau der Befunde zeigt, dass Studien, in denen latente Variablen modelliert werden, häufiger einen Faktor (Ackerman et al., 2002) oder zwei hoch korrelierte Faktoren finden (Kane et al., 2004; Süß et al., 2002), während korrelative Studien tendenziell die Domänenspezifitäts-Annahme stützen (Vock, 2005). Für diese widersprüchliche Befundlage werden verschiedene Gründe angeführt, wie beispielsweise die Verwendung zu homogener bzw. zu kleiner Stichproben (Miyake, 2001) oder die Vernachlässigung visuell-räumlicher gegenüber verbal-numerischen Aufgaben. Doch abgesehen von der Problematik methodischer Artefakte sprechen sich viele Forscher für die Betrachtung von WM als sowohl domänenspezifische, als auch –generelle Instanz aus (Hitch, Towse & Hutton, 2001; Kintsch, Healy, Hegarty, Pennington & Salthouse, 1999; Miyake, 2001), die vermutlich ähnlich wie Intelligenz eine hierarchische Organisationsstruktur aufweist.

1.1.4 Entwicklung des Arbeitsgedächtnisses im Kindesalter

Ein Großteil der Untersuchungen und Experimente, die zum heutigen Verständnis des Konzepts WM beigetragen haben, wurden an Erwachsenen durchgeführt. Für Studien wie die hier vorliegende ist jedoch die Frage, inwiefern Modellvorstellungen und Befunde von Erwachsenen auf Kinder generalisierbar sind, hochrelevant. So wäre es beispielsweise denkbar, dass sich die von Baddeley postulierte Multi-Komponenten-Struktur erst in der Adoleszenz manifestiert und im frühen Kindesalter nur wenig ausdifferenziert ist. Daher werden nun die wichtigsten Erkenntnisse bezüglich Struktur und Veränderung des WMs im Kindesalter zusammengefasst.

Allgemein zeigt sich, dass die Leistung des KZGs ab dem Vorschulalter bis hin zur Adoleszenz stetig ansteigt (Case, Kurland & Goldberg, 1982; Dempster, 1985; Hulme, Thomson, Muir & Lawrence, 1984; Isaacs & Vargha-Khadem, 1989; Siegel, 1994). Gathercole, Pickering, Ambridge und Wearing (2004) konnten diesen linearen Anstieg auch für alle von Baddeley postulierten Komponenten des WMs getrennt nachweisen. So herrscht Konsens darüber, dass im Laufe der Kindheit auch die Arbeitsgedächtniskapazität allgemein signifikant ansteigt (Cowan et al., 2005; Engel de Abreu, Conway & Gathercole, 2010; Gathercole, 1999). Allerdings unterliegen ZE, PL und VSSP auch spezifischen Veränderungen. Wie in Kapitel 1.1.2.1 erläutert, besteht die PL eines Erwachsenen aus einem verbalen Speicher, der durch einen Rehearsalmechanismus unterstützt wird. Dieser ist jedoch bei Kindern unter sieben Jahren noch nicht, bzw. nur rudimentär ausgebildet, so dass die Leistungssteigerung der PL mit der Entwicklung dieser Rehearsalfähigkeit in Verbindung gebracht wird (siehe Gathercole & Hitch, 1993). Durch Rehearsal können größere Mengen verbalen Inputs im Speicher aufrecht erhalten werden (Hulme, Thomson, Muir & Lawrence, 1984). Cowan et al. (1992, 1998) weisen außerdem darauf hin, dass Informationen mit steigendem Alter schneller gescannt werden können, was ebenfalls die Leistung positiv beeinflusst.

Auch die Verarbeitung von visuellem Input erfährt entwicklungsbedingte Veränderungen, die auf Testebene in höheren Leistungen münden. Hier spielt die Altersgrenze von sieben Jahren ebenfalls eine wichtige Rolle. Während Kinder unterhalb dieser Grenze zur Verarbeitung visueller Information ausschließlich auf den VSSP zurückgreifen, sind ältere Kinder zunehmend in der Lage visuelle Stimuli zu rekodieren und für die PL nutzbar zu machen (Gathercole et al., 2004). Allerdings

sind Leistungssteigerungen auch festzustellen, wenn nicht rekodierbares Stimulusmaterial zum Einsatz kommt. Eindeutige Erklärungen gibt es für diese Befunde bisher nicht (Pickering, Gathercole, Hall & Llyod, 2001). Neben einer allgemeinen Kapazitätssteigerung des VSSP per se (Logie & Pearson, 1997) werden sowohl die Rolle des Vorwissens diskutiert, also auch der effizientere Einsatz von Strategien und eine stärkere Einbindung der ZE (siehe Gathercole, 2001).

Untersuchungen mit Fokus auf die Organisation des Arbeitsgedächtnisses im frühen Kindesalter zeigen, dass sich die von Baddeley postulierte Mehr-Komponenten-Struktur bestehend aus ZE, PL und VSSP bereits im Alter von sechs Jahren etabliert hat (Gathercole, Pickering, Ambridge & Wearing, 2004). Bezüglich der Beziehung zwischen den einzelnen Komponenten scheint es wenig entwicklungsbedingte Veränderungen zu geben. Während sich PL und VSSP in verschiedenen Altersgruppen als separierbar und relativ unabhängig voneinander erweisen, ist die Beziehung der Subkomponenten zur ZE recht deutlich (Gathercole et al. 2004; Jarvis & Gathercole, 2003). Bereits in der Altersgruppe von Fünf- bis Achtjährigen konnte die Unabhängigkeit zwischen PL und VSSP nachgewiesen werden (Pickering, Gathercole & Peaker, 1998). In einer Stichprobe von sechs- bis siebenjährigen Kindern konnten Gathercole und Pickering (2000a) außerdem zeigen, dass die ZE mit der PL zwar moderat assoziiert war, beide Komponenten jedoch wie bei Erwachsenen separiert werden konnten. Allerdings war es in dieser Studie nicht möglich, den VSSP von der ZE abzuspalten. Auch Jarvis und Gathercole (2003) wiesen bei Elf- bis Vierzehnjährigen voneinander unabhängige verbale und visuell-räumliche Gedächtniskomponenten nach. Diese Befunde bestätigen einerseits die Baddeley'sche Modellvorstellung und decken sich außerdem mit Ergebnissen der neuroanatomischen Forschung, welche die Separierbarkeit verbaler und visuell-räumlicher Systeme im Gehirn belegen (Smith & Jonides, 1997; Vallar & Papagno, 2002).

Uneinigkeit herrscht weiterhin bezüglich der die ZE betreffenden Veränderungsprozesse. Baddeley selbst bewertet die ZE zwar als wichtigste, jedoch gleichzeitig als die am wenigsten verstandene und untersuchte Komponente (Baddeley, 2006). Wie bereits in Kapitel 1.1.2.1 erwähnt, werden typischerweise Storage-and-Processing-Aufgaben (SAPTs) eingesetzt, um die Kapazität der ZE abzugreifen. Dabei deckt die Processing-Komponente vorrangig die Leistung der ZE ab, während die Storage-Komponente die Leistung der PL bzw. des VSSP widerspiegelt (Baddeley & Logie, 1999; Duff & Logie, 2001; LaPointe & Engle, 1990).

Doch welche Entwicklungsprozesse nun die Leistungssteigerung in diesen Aufgaben bedingen, wird nach wie vor debattiert. Verfechter der *Resource-Sharing-Hypothese* gehen davon aus, dass im Arbeitsgedächtnissystem ein allgemeines Kapazitätslimit existiert, wodurch sowohl die Speicher- als auch die Prozessierfähigkeit in SAPTs begrenzt wird. Damit wird dieses Modell zur Prüfung der Domänengeneralitäts-Hypothese genutzt. Dem Modell folgend steigern sich die Leistungen in SAPTs nicht durch eine Ausweitung der verfügbaren Ressourcen. Vielmehr wird die Leistungssteigerung damit erklärt, dass SAPTs mit zunehmendem Alter weniger prozessbeanspruchend werden, wodurch mehr Ressourcen zur Informationsspeicherung freigesetzt werden (Case, Kurland & Goldberg, 1982). Als Kritik an diesem Modell wird angeführt, dass der postulierte Trade-off zwischen Speicherung und Prozessieren bisher nicht eindeutig nachgewiesen werden konnte (Gathercole & Pickering, 2000a). Eine alternative Erklärung liefert die *Task-Switching-Hypothese*, die besagt, dass bei der Lösung komplexer Spannaufgaben zwischen Verarbeitungs- und Speicherkomponente hin- und her gewechselt wird. Speichern und Prozessieren werden als zu zwei getrennten Systemen zugehörige Prozesse betrachtet. Es wird also entweder gespeichert oder prozessiert. Je effizienter nun verarbeitet wird, umso kürzer muss die Information gehalten werden, wodurch Vergessen unwahrscheinlicher wird und die Leistung ansteigt (Hitch et al., 2001; Towse et al., 1998). Vermutlich wird die Leistung in Spannaufgaben jedoch stets durch multiple Prozesse determiniert, so dass es zum jetzigen Zeitpunkt sinnvoll erscheint, beide Hypothesen als sich ergänzend und nicht als sich wechselseitig ausschließend zu betrachten. Auch Swanson (2008) postuliert, dass die entwicklungsbedingten Änderungen der Arbeitsgedächtnisleistung sowohl auf Veränderungen von Prozessen beruhen, wie der Fähigkeit zur Aufmerksamkeitskontrolle, als auch auf Vorgängen, die die Speicherfunktion betreffen.

Neuere Studien beschäftigen sich außerdem mit der Differenzierung der verschiedenen Aspekte exekutiver Funktionen. Eine diesbezüglich wichtige Arbeit stammt von Miyake et al. (2000), welche die drei Prozesse *Shifting*, *Updating* und *Inhibition*² als Kernbereiche exekutiver Funktionen identifizieren konnten. Es

² Der Prozess des Shifting meint die Fähigkeit zwischen Aufgaben, Strategien oder Operationen hin und her wechseln zu können. Inhibition bedeutet die Unterdrückung von bestimmten Verhaltenstendenzen zu Gunsten von Verhalten, das im Sinne einer Zielerreichung angemessener ist. Updating steht für die Überwachung und Bewertung neu eingehender Information hinsichtlich der Aufgabenrelevanz bei nachfolgender Änderung der im Gedächtnis gehaltenen Information (St. Clair-Thompson & Gathercole, 2006).

existieren zwar Hinweise, dass sich diese Prozesse auch bei Kindern als trennbare Funktionen nachweisen lassen, allerdings ist die Befundlage hinsichtlich der Anzahl und Eigenschaften dieser Komponenten uneindeutig (Lehto, Juujarvi, Kooistra & Pulkkinen, 2003; Levin et al., 1996; Welsh, Pennington & Groisser, 1991). In einer neueren Studie konnten St. Clair-Thompson und Gathercole (2006) bei Kindern zwischen elf und zwölf Jahren zwar zwischen Updating und Inhibition unterscheiden, allerdings konnte Shifting nicht als eigenständiger Prozess nachgewiesen werden. Dieser Befund könnte auf entwicklungsbedingte Organisationsunterschiede des Systems exekutiver Funktionen zwischen Kindern und Erwachsenen hindeuten. So wiesen bereits Senn, Espy und Kaufmann (2004) darauf hin, dass die geistige Flexibilität im Kindesalter möglicherweise weniger leicht von anderen Prozessen zu trennen ist als bei Erwachsenen. Doch es ist zu erwähnen, dass die Reliabilität der Shifting-Aufgaben unbekannt ist, so dass den Befunden auch methodische Ursachen zu Grunde liegen könnten (St. Clair-Thompson & Gathercole, 2006).

Exekutive Funktionen und Aufmerksamkeit stehen in enger Beziehung zueinander. Aufmerksamkeit ermöglicht es, sich auf relevante Informationen zu konzentrieren (Engle et al., 1999) bzw. Irrelevantes zu unterdrücken (Inhibition) (Lustig, May & Hasher, 2001; Unsworth & Engle, 2007). Es ist naheliegend, dass Aufmerksamkeit, bzw. die Fähigkeit zur Kontrolle dieser, ebenfalls entwicklungsbedingten Veränderungen unterliegt, wodurch Leistungsunterschiede bzw. altersbedingte Leistungszuwächse in Arbeitsgedächtnisaufgaben mit verursacht werden könnten. So lieferte beispielsweise die Studie von Gomes, Molholm, Christodoulou, Ritter und Cowan (2000) Hinweise darauf, dass die Fähigkeit der Aufmerksamkeitskontrolle im Kindesalter deutlich geringer ausgeprägt ist als im Erwachsenenalter.

Schließlich kann die Erforschung der Separierbarkeit von Arbeits- und Kurzzeitgedächtnis im Kindesalter als aktuelle Forschungsfrage genannt werden. Diese Frage ist für Kinder ebenso wenig abschließend beantwortet wie für Erwachsene. Einerseits betrachten einige Autoren die Unterscheidung von WM und KZG als schwierig, da die eingesetzten Aufgaben nie reine WM- oder KZG-Indikatoren darstellen und es stets eine Konfundierung gibt. Schließlich entsprechen sich WM und KZG nach Engle et al. (1999) weitestgehend, mit dem Zusatz, dass WM zusätzlich den Aspekt der kontrollierenden Aufmerksamkeit umfasst. Eine Studie an acht- bis elfjährigen Kindern bestätigte diese Annahme, da Arbeits- und Kurzzeitgedächtnisaufgaben auf einem gemeinsamen Faktor luden, welcher mittels einer Hauptkomponentenanalyse extrahiert worden war (Hutton & Towse, 2001).

Doch andererseits konnten in einer aktuellen Studie von Engel de Abreu et al. (2010), in der eine Stichprobe von fünf- bis neunjährigen Kindern untersucht wurde, unterscheidbare Komponenten für KZG und WM auf latenter Ebene nachgewiesen werden. Auch die Ergebnisse der Studie von Swanson (2008) an Grundschulkindern sprachen für die Separierbarkeit von WM und KZG im frühen Kindesalter. Denn obwohl eine Korrelation von $r = .55$ zwischen den zwei latenten Gedächtnisfaktoren auftrat, luden die Kurzzeitgedächtnis- bzw. Arbeitsgedächtnisaufgaben auf zwei separierbaren Faktoren.

Zusammengefasst zeigen die dargelegten Befunde, dass zahlreiche Belege dafür existieren, dass sich die grundlegende Struktur der Instanz WM bereits im frühen Kindesalter ausbildet und keine weitreichenden strukturellen Änderungen im Laufe des Heranwachsens erfährt. Dennoch sind weitere Studien zur Organisation des WMs wünschenswert, um beispielsweise die Frage nach der Unterscheidbarkeit von WM und KZG beantworten zu können oder die Beziehung zwischen WM und Aufmerksamkeit intensiver zu beleuchten.

1.1.5 Zusammenfassende Betrachtung

Trotz der angesprochenen Debatten um die dem Arbeitsgedächtnissystem zugrunde liegenden Prozesse und Strukturen, gibt es im Kern über alle Konzeptionen hinweg einen Konsens die folgenden grundlegenden Punkte betreffend.

Die Vorstellung, dass sich WM in einem eng umschreibbaren Hirnareal lokalisieren lässt, welches für die Speicherung von Information zuständig ist, hat sich als nicht haltbar erwiesen, weder auf kognitiver noch auf neuronaler Ebene. So verdeutlichen zahlreiche Befunde, dass die Instanz WM eine Vielzahl verschiedener Prozesse umfasst, die die Beteiligung unterschiedlicher Hirnareale erfordern, wobei besonders der PFC eine wichtige Rolle zu spielen scheint (Miyake & Shah, 1999). Diesem Verständnis entsprechend vereint der Begriff WM alle Prozesse und Mechanismen, die kontrollierend oder regulierend an der Lösung kognitiver Aufgaben beteiligt sind, oder die für die Aufrechterhaltung von zur Aufgabenbewältigung notwendiger Information verantwortlich sind. Das System WM arbeitet dabei mit Hilfe verschiedener Repräsentationen und Codes und umfasst mehrere Subsysteme, was der Idee einer unitären Instanz widerspricht. Arbeitsgedächtnis und Langzeitgedächtnis sind eng verwoben, schließlich können viele

Arbeitsgedächtnisinhalte als aktuell aktivierte Langzeitrepräsentation verstanden werden (Miyake & Shah, 1999).

Trotz der Bemühungen um Integration der verschiedenen Ansätze existieren bis zum heutigen Zeitpunkt weitreichende ungeklärte Fragen, die weiterer Forschung bedürfen. Es mangelt beispielsweise an einer exakten Ausformulierung der unterschiedlichen Funktionen, die der exekutiven Kontrollinstanz obliegen, sowie der Beziehung dieser verschiedenen Funktionen zueinander. Ein weiterer Diskussionspunkt betrifft die Binnenstruktur des Arbeitsgedächtnissystems. Auch wenn die Tatsache unumstritten ist, dass WM zu Teilen domänen-spezifisch organisiert ist, bleibt offen, bis zu welchem Grad und welche verschiedenen Quellen zu dieser Organisationsform beitragen. So werden nicht nur spezifische Subsysteme diskutiert, es wird auch darüber debattiert, inwiefern Verarbeitungsgeschwindigkeit und kognitive Strategien spezifischer Natur sein könnten (Miyake & Shah, 1999).

Erwähnenswert ist schließlich auch die ungeklärte Beziehung zwischen WM, Bewusstsein und Aufmerksamkeit. Dass zwischen diesen Konstrukten ein enger Zusammenhang besteht, wird in verschiedenen Modellen explizit ausformuliert, doch wenig beforscht ist bisher die Natur dieses Zusammenspiels (Miyake & Shah, 1999).

Es ist zu erwarten, dass durch die zukünftige Forschung und die Auseinandersetzung mit diesen und anderen offenen Fragen die bestehenden Arbeitsgedächtnismodelle weiter spezifiziert werden. Die Weiterentwicklung eines integrativen Ansatzes bezüglich Organisation und Funktionsweise des WMs sollte so stetig voran getrieben werden.

1.2 Intelligenz

Menschen unterscheiden sich hinsichtlich ihrer kognitiven Fähigkeiten. Somit stellt die Intelligenzforschung ein zentrales Forschungsgebiet innerhalb der differentiellen Psychologie dar. Gottfredson und Saklofske (2009) messen ihr gar den größten Stellenwert innerhalb der gesamten Psychologie bei. Auch in der empirischen Bildungsforschung nimmt die Intelligenzforschung einen zentralen Stellenwert ein. Denn die Frage nach den Bedingungsfaktoren schulischer Leistung kann nicht ohne die Berücksichtigung der Intelligenz erfolgen, da diese Lernprozesse erleichtert, wodurch es intelligenteren Kindern einfacher fällt, sich Wissen und Kompetenzen in der Schule anzueignen (Helmke & Schrader, 2006).

Bevor jedoch auf die Zusammenhänge zwischen Intelligenz und Schulerfolg eingegangen wird (Kapitel 1.5.1), erfolgt zunächst ein theoretischer Überblick über das Konzept Intelligenz. Nach einer kurzen Einführung in das Themengebiet der Intelligenzforschung (Kapitel 1.2.1), erfolgt anschließend eine Begriffsbestimmung (1.2.2), bevor schließlich auf die Struktur der Intelligenz (Kapitel 1.2.3), sowie die Bedeutung des *g*-Faktors näher eingegangen wird (Kapitel 1.2.4). Nachfolgend wird die Stabilität der Intelligenz näher betrachtet (Kapitel 1.2.5), bevor der Zusammenhang zwischen Intelligenz und Arbeitsgedächtnis dargestellt wird (Kapitel 1.2.6).

1.2.1 Überblick über das Themengebiet der Intelligenzforschung

Die Ursprünge dieser Forschungstradition reichen weit zurück und so verwundert es nicht, dass Intelligenz als das meist beforschte psychologische Konzept gilt (Gottfredson & Saklofske, 2009).

Als Meilenstein der Intelligenzforschung gilt Spearman's Werk „General Intelligence, objectively determined and measured“, welches bereits 1904 erschien und sich als eine der ersten Arbeiten mit dem *g*-Faktor bzw. der allgemeinen Intelligenz befasste (Spearman, 1904). Als „eigentlicher Begründer der Intelligenzforschung“ (Mehlhorn & Mehlhorn, 1981, S. 20) kann Alfred Binet verstanden werden, der zusammen mit Simon einen ersten Intelligenztest entwickelte (Binet & Simon, 1905a; Binet & Simon, 1905b; Binet & Simon, 1905c). Die in

Frankreich eingeführte Verpflichtung zur Vorlage eines medizinisch-pädagogischen Gutachtens im Falle der Versetzung eines Kindes von einer Normal- auf eine Sonderschule, erforderte die Entwicklung eines wissenschaftlich fundierten Instruments, das es ermöglichte, anhand der Testergebnisse interindividuelle Unterschiede zwischen Schulkindern messbar zu machen, um Entwicklungsrückstände rechtzeitig aufdecken zu können. Binet und Simon entwickelten das Konzept des Intelligenzalters, um in einer Normalpopulation minderbegabte Kinder zu identifizieren (Gottfredson & Saklofske, 2009). Diese Idee wurde 1912 von Stern weiterentwickelt. Durch die Einführung des Intelligenzquotienten, welcher Intelligenzalter und Lebensalter ins Verhältnis setzt, wurde der Vergleich von Personen unterschiedlichen Alters ermöglicht (Mehlhorn & Mehlhorn, 1981). Das Zeitalter der Industrialisierung und die damit verbundene Notwendigkeit, Jugendliche für die unterschiedlichen Anforderungen industrieller Berufe auszubilden, führte außerdem zu einem enormen gesellschaftlichen und wissenschaftlichen Interesse an der Erforschung menschlicher Begabungsunterschiede (Mehlhorn & Mehlhorn, 1981). Schnell stand die praktische Relevanz der Tests im Vordergrund. So wurden diese beispielsweise zur Zeit des ersten Weltkrieges in den USA als schnelles und kostengünstiges Screening-Verfahren zu Rekrutierungszwecken eingesetzt (Gottfredson & Saklofske, 2009). Firmen wurden gegründet, die sich mit der Entwicklung neuer Testverfahren beschäftigten, um beispielsweise Personalauswahlverfahren effizienter und fairer zu gestalten. Daraus resultierend existierte bereits in den dreißiger Jahren des zwanzigsten Jahrhunderts eine Fülle an Verfahren (Gottfredson & Saklofske, 2009). Heute basieren die meisten modernen Verfahren auf diesem psychometrischen Ansatz, der der Messung interindividueller Differenzen in kognitiven Fähigkeiten mit Hilfe genormter Verfahren dient (Gustafsson & Undheim, 1996). Auch die vorliegende Arbeit folgt diesem Ansatz.

Zwischenzeitlich blickt die Intelligenzforschung also auf gut 100 Jahre Forschungsarbeit zurück. Besonders die Intelligenzdiagnostik hat sich als Forschungsschwerpunkt etabliert und „gehört heute zu den am weitesten entwickelten Bereichen der bürgerlichen Psychologie“ (Mehlhorn & Mehlhorn, 1981, S. 63). Auch die Erforschung der Struktur der menschlichen Intelligenz oder der Beziehung zu anderen kognitiven Fähigkeiten stellen nach wie vor aktuelle Themengebiete dar.

Als vielversprechende Erweiterung der klassischen Intelligenzforschung sind die differentiellen Neurowissenschaften zu nennen, deren Ziel es ist, mittels verhaltensgenetischer oder bildgebender Verfahren die biologischen Grundlagen der menschlichen Intelligenz zu erforschen (Deary, Penke & Johnson, 2010).

Bevor die Struktur der Intelligenz näher beleuchtet wird, wird zunächst auf den Intelligenzbegriff genauer eingegangen.

1.2.2 Zum Begriff der allgemeinen Intelligenz

Ähnlich wie in der Arbeitsgedächtnisforschung, herrscht auch in der Intelligenzforschung eine gewisse Uneinigkeit hinsichtlich Definition und Verwendung des Intelligenzbegriffs (Süß, 2001). Für Brocke und Beauducel (2001) stellt Intelligenz ein komplexes Konstrukt dar, welches vielfältige kognitive Teilfähigkeiten umfasst. Hofstätter (1957) versteht unter Intelligenz die Menge an Fähigkeiten, die die innerhalb einer Kultur erfolgreichen Individuen teilen. Und Sternberg (1997) betont, dass intelligentes Verhalten die Fähigkeit zur Anpassung bzw. Gestaltung der Umwelt des Menschen meint.

Nicht zuletzt aufgrund der Fülle an Definitionen gilt Intelligenz als das in der öffentlichen Diskussion umstrittenste psychologische Phänomen (Gottfredson & Saklofske, 2009). Doch zwei umfassende Expertenbefragungen in den Jahren 1921 sowie 1981 zeigten, dass sich trotz einer Vielzahl an verbalen Definitionen ein weitreichender Konsens hinsichtlich elementarer Kennzeichen menschlicher Intelligenz etabliert hat. So stimmten 52 Intelligenzforscher folgender breit gefasster Definition zu:

„Intelligence is a very general capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings-`catching on´,`making sense´ of things, or `figuring out´ what to do. Intelligence, so defined, can be measured, and intelligence tests measure it well” (Gottfredson, 1997a, S. 13).

Die in dieser Expertenbeschreibung genannten Attribute wie abstraktes Denken, Entscheidungsfindung, oder Problemlösen können zusammengefasst als psychometrische Definition der allgemeinen kognitiven Fähigkeit bezeichnet werden (Sternberg & Detterman, 1986; Süß, 2001).

Trotz eines gewissen „common sense“ hinsichtlich des grundlegenden Verständnisses des Intelligenzbegriffs, sind die je nach Forschungstradition vorliegenden Unterschiede im Hinblick auf Annahmen, die Intelligenzstruktur bzw. die Bedeutung einzelner Komponenten betreffend, nicht von der Hand zu weisen (Jensen, 1998). Dies spiegelt sich auch in der Entwicklung verschiedener Strukturmodelle wider, welche im folgenden Kapitel thematisiert werden.

1.2.3 Strukturmodelle

Im Kern der Diskussion um die Intelligenzstruktur steht die Frage, ob menschliche Intelligenz eine generelle Fähigkeit darstellt, oder ob unterscheidbare, unabhängige Teil-Fähigkeiten existieren.

Charles Spearman (1904) unterschied in seinem *Zwei-Faktoren-Modell* spezifische Faktoren (s- Faktoren) von einem Generalfaktor zweiter Ordnung, auch *g* genannt (general factor). Ausgehend von der Betrachtung der korrelativen Zusammenhänge verschiedener Intelligenzaufgaben schlussfolgerte er, dass die Leistungen in diesen Tests stets durch zwei Komponenten beeinflusst werden: die für die jeweiligen Aufgaben *spezifischen Anteile* (s) und eine *allgemeine Fähigkeit* (g), die einen Einfluss auf sämtliche kognitiven Leistungen ausübt (Mehlhorn & Mehlhorn, 1981). Für den Zusammenhang zwischen zwei Aufgaben ist somit laut Spearman der gemeinsame Aspekt – also *g* - verantwortlich. Der Befund, dass alle Intelligenztests stets positiv korrelieren, wurde vielfach repliziert und kann als Bestätigung der Existenz eines *g*-Faktors gesehen werden. Dies bedeutet, dass im Allgemeinen Personen, die in einer bestimmten Domäne, wie beispielsweise *Reasoning*, gut abschneiden, dies auch in anderen Bereichen tun (Deary et al., 2010). Dennoch sprechen einige Befunde gegen Spearman's Modell, wie etwa die Tatsache, dass sich die s-Faktoren nicht als unkorreliert erwiesen haben (Holling, Preckel & Vock, 2004) und somit die Korrelationen zwischen Tests nicht vollständig durch *g* zu erklären sind.

Konträre Modelle vertraten Thorndike (1924), der von mehreren Einzelfaktoren ausging bzw. Thurstone (1938), dessen *Primärfaktoren-Modell* sieben unabhängige und gleichberechtigte Faktoren³ postulierte. Er ging davon aus, dass die Lösung einer Aufgabe immer die Beteiligung mehrerer Gruppenfaktoren erfordert, allerdings mit unterschiedlicher Gewichtung. Doch zwischenzeitlich gilt diese Theorie als widerlegt, denn es konnte gezeigt werden, dass sich selbst aus Thurstone's eigenen Daten ein *g*-Faktor ableiten lässt (Deary et al., 2010).

Eine Weiterentwicklung der frühen Modelle von Spearman und Thurstone stellt das *Modell der fluiden und kristallinen Intelligenz* von Horn und Cattell (1966) dar, welches ebenfalls eine hierarchische Struktur postuliert und als eines der einflussreichsten Strukturmodelle zu bewerten ist. Besonders den Faktoren *fluide Intelligenz* (*gf*) und *kristalline Intelligenz* (*gc*), welche auf einer mittleren Ebene angesiedelt sind, kommt eine große Bedeutung zu, weshalb auf diese näher eingegangen wird. Während sich *gf* auf die Fähigkeit bezieht, neue Probleme und Situationen ohne die maßgebliche Beteiligung vorheriger Lernerfahrung zu bewältigen, spiegelt *gc* die sich kumulierenden Effekte von Lernerfahrung und faktischem Wissen wider. Damit gilt *gf* als angeborene, relativ kulturunabhängige Grundlage der kristallinen Intelligenz, weshalb Cattell's Theorie auch mit dem Begriff der *Investmenttheorie* verknüpft ist. So ist Wissen nach Cattell das Resultat investierter Intelligenz. Nach Horn (1968) spiegelt *gf* besonders Fähigkeiten wie *Induktives Denken*, *Figurale Beziehungen* oder *Gedächtnisspanne* wider und wird üblicherweise durch nonverbale Aufgaben erfasst, die das Erkennen und Anwenden von Regeln erfordern. Kristalline Intelligenz hingegen ist deutlich geprägt durch kulturabhängige Einflüsse und deshalb stark mit verbalem Verständnis oder semantischem Schlussfolgern assoziiert. Für eine Unterscheidung dieser beiden Intelligenzfaktoren sprechen besonders die unterschiedlichen Entwicklungsverläufe über die Lebensspanne hinweg. So erreicht *gf* frühzeitiger ihren Gipfel und fällt ab einem Alter von ca. 22 Jahren kontinuierlich ab, während *gc* bis ins hohe Alter anwachsen kann (Horn & Cattell, 1967; Mehlhorn & Mehlhorn, 1981). Beide Dimensionen korrelieren mit etwa $r = .50$, allerdings lädt *gf* höher auf *g* als *gc* (Brody, 1992).

³ Mit Hilfe der multiplen Faktorenanalyse extrahierte Thurstone die Primärfaktoren *Merkfähigkeit* (memory), *Verbales Verständnis* (verbal comprehension), *Wortflüssigkeit* (word fluency), *Rechenfähigkeit* (number), *Räumliches Vorstellungsvermögen* (space), *Wahrnehmungsgeschwindigkeit* (perceptual space) sowie *Schlussfolgerndes Denken bzw. Erkennen von Regeln* (reasoning).

Insgesamt gilt Cattells Modell auch heute noch als heuristisch wertvoll.

Aktuelle Intelligenzmodelle wie beispielsweise das von Jäger (1982) konzipierte *Berliner Intelligenzstrukturmodell* (BIS-Modell), versuchen die oben dargestellten Konzepte unter Berücksichtigung der Vielschichtigkeit kognitiver Fähigkeiten zu integrieren. Als der „bislang umfassendste Entwurf der Intelligenzstrukturforschung“ (Süß, 2001) wird das *Three-Stratum-Modell* von Carroll (1993) bezeichnet. Dieser reanalysierte mittels faktorenanalytischer Methoden über 460 Datensätze, um zu einer Neuordnung der Konstrukte bzw. einer Integration der divergierenden Befunde und Konzeptionen zu gelangen. Die Analyse der Datensätze mit heterogenen Aufgabentypen lieferte eine hierarchische Struktur bestehend aus drei Schichten (Strata) unterschiedlicher Generalität. Besonders hervorzuheben ist der deutliche Nachweis des *g*-Faktors im Sinne der allgemeinen Intelligenz, welcher in seinem Modell auf Stratum III den übergeordneten Faktor höchster Generalität darstellt. Auf der Ebene unterhalb des *g*-Faktors (Stratum II) differenziert Carroll zwischen acht Subkomponenten mittlerer Generalität, welche durch *g* beeinflusst werden: *Fluide und Kristalline Intelligenz, Allgemeines Gedächtnis und Lernen, Visuelle und Auditive Wahrnehmung, Kognitive Geschwindigkeit, sowie Abruffähigkeit und Verarbeitungsgeschwindigkeit* (Brocke & Beauducel, 2001; Süß, 2001). Jedem dieser Second-Order-Faktoren werden auf Stratum I zwischen vier und fünfzehn Primärfaktoren zugeordnet, so dass sich auf der untersten Ebene insgesamt 68 spezifische kognitive Fähigkeiten unterscheiden lassen, auf welche die höheren Ebenen einen Einfluss nehmen.

Problematisch an der Konzeption Carrolls ist, dass bislang keine Testbatterie existiert, die auf seiner Modellvorstellung beruht, so dass der empirische Nachweis des Modells anhand eines einzelnen Datensatzes aussteht (Süß, 2001). Außerdem wird das methodisch uneinheitliche Vorgehen Carrolls bei der Auswertung bemängelt, insbesondere im Hinblick auf die Anwendung verschiedener Rotationstechniken (Brocke & Beauducel, 2001). Dennoch konnte das von Carroll postulierte Modell mit Hilfe konfirmatorischer Faktorenanalysen empirisch untermauert werden (Bickley, Keith & Wolfle, 1995) und stellt für viele die bislang beste Repräsentation kognitiver Fähigkeiten dar (Gottfredson & Saklofske, 2009).

McGrew (2005) wies darauf hin, dass sich das Cattell'sche *gf-gc-Modell* und Carrolls Befunde integrieren lassen, was zur Entwicklung des *C-H-C-Modells* (Cattell-Horn-Carroll-Modell) führte, welches in der Forschung derzeit diskutiert wird. Dies spiegelt

sich auch in der Berücksichtigung dieser Konzeption in aktuellen Testrevisionen wider, beispielsweise im HAWIK-IV (Hamburg-Wechsel-Intelligenztest für Kinder; Petermann & Petermann, 2008).

Die zusammenfassende Betrachtung der unterschiedlichen Strukturmodelle der Intelligenz ergibt zwar eine gewisse Variation hinsichtlich Anzahl und Inhalt der jeweiligen Modellkomponenten. Dennoch kristallisiert sich die Robustheit des Befundes einer hierarchischen Intelligenzstruktur mit einem an der Spitze stehenden Generalfaktor und spezifischeren Fähigkeiten heraus, die auf einer oder mehreren Ebenen unterhalb des *g*-Faktors angesiedelt sind (Vock, 2005). Dieses hierarchische Modell lässt sich auf alle bisher untersuchten Altersgruppen, Geschlechter und Kulturen übertragen (Jensen, 1998).

Die Theorien von Autoren, die die Existenz multipler Intelligenzen postulieren (Gardner, 1983), welche unkorrelierte Fähigkeiten darstellen sollen, haben sich im Gegenzug als nicht haltbar erwiesen.

Auch wenn zahlreiche Studien die Existenz eines *g*-Faktors konsistent belegen, sagt der bloße Nachweis noch nichts über die eigentliche Bedeutung dieses Phänomens aus. Deshalb soll darauf im Folgenden näher eingegangen werden.

1.2.4 Zur Bedeutung des *g*-Faktors

Das Konzept des *g*-Faktors kognitiver Fähigkeiten kann als das meist akzeptierte der gegenwärtigen Intelligenzstrukturforschung angesehen werden (Neisser et al., 1996). Dabei gilt die positive Interkorrelation unterschiedlicher Tests zur Messung kognitiver Fähigkeiten (positive Mannigfaltigkeit) als notwendige und hinreichende Bedingung der Ableitung eines *g*-Faktors (Humphreys & Stark, 2002; Jensen, 1993). Korreliert man die aus verschiedenen Testbatterien extrahierten Generalfaktoren, ergeben sich Zusammenhänge von über $r = .90$, was als weiterer Beleg der Existenz eines *g*-Faktors gewertet werden kann (Jensen, 1998). Zwar weisen Humphreys und Stark (2002) darauf hin, dass *g* nicht direkt beobachtbar ist und warnen vor der Interpretation dieses phänotypischen Merkmals als rein biologische Kapazität. Dennoch wird *g* eine über die faktorenanalytische Bedeutung hinausgehende inhaltliche Relevanz zugesprochen, im Sinne eines „common core to all tested

mental abilities" (Gottfredson & Saklofske, 2009, S.188), der in unterschiedlichem Maße alle kognitiven Fähigkeiten bzw. Leistungen bedingt (Gottfredson, 1997b). Empirisch gilt die Existenz des *g*-Faktors als gesichert (Deary, 2001; Gottfredson 1997a, 1997b; Plomin, DeFries, McClearn & McGuffin, 2008).

Hervorzuheben ist die immense Bedeutung des *g*-Faktors als „real-life“-Kriterium, denn die kognitive Fähigkeit eines Menschen beeinflusst nahezu jeden Bereich des Lebens. So nimmt *g* nicht nur einen Einfluss auf den Bildungsverlauf, was Korrelationen zu Berufs-, Studien- oder Schulerfolg belegen (Boulanger, 1981; Fleming & Malone, 1983; Giesen, Gold, Hummer & Jansen, 1986; Schmidt-Atzert & Deter, 1993; Schmitt, Gooding, Noe & Kirsch, 1984; Steinkamp & Maehr, 1983). Auch der sozioökonomische Status korreliert signifikant in einem Bereich zwischen $r = .30$ und $.40$ mit *g* (Brody, 1992). Ein niedriger IQ erhöht außerdem das Risiko für Armut im Erwachsenenalter (Gottfredson & Saklofske, 2009), Schulabbruch (ebd.), Kriminalität (Neisser et al., 1996) und Arbeitslosigkeit (Gottfredson, 1998). Darüber hinaus bestätigen zahlreiche Befunde aus dem Bereich der Gesundheitspsychologie die weitreichende Relevanz der allgemeinen kognitiven Fähigkeit. So konnte nachgewiesen werden, dass für Menschen mit höherer allgemeiner Intelligenz die Wahrscheinlichkeit jung zu sterben geringer ausfiel und sie außerdem im mittleren und hohen Alter einen besseren gesundheitlichen Zustand aufwiesen (Batty, Deary, & Gottfredson, 2007). Außerdem konnte eine Beziehung zur Wahrscheinlichkeit, eine chronische Erkrankung zu erleiden, gefunden werden (Gottfredson & Saklofske, 2009).

Zusammenfassend stellt Gottfredson (2002) fest, dass zwar kein Lebensbereich einzig und allein durch den *g*-Faktor determiniert wird, umgekehrt ist er offensichtlich jedoch auch für keinen Bereich irrelevant. Außerdem wurde bis dato kein alternativer Faktor gefunden, der den *g*-Faktor hinsichtlich seines praktischen Nutzens und der Generalität zu übertreffen vermag. Auch Deary et al. (2010) betonen die Bedeutung von *g*: „it accounts for a substantial large amount of variance, it is the source of most of the predictive power of cognitive tests and, as discussed below, it is the locus of most of the genetic variance" (S. 202).

Es ist anzumerken, dass der in dieser Studie eingesetzte Intelligenztest CFT 20-R (Culture Fair Test 20-R; Weiß, 2006) streng genommen keinen breiten *g*-Faktor misst, sondern die Gundintelligenz im Sinne der *General Fluid Ability* nach Cattell (1963). Da jedoch *gf* als bester Indikator für *g* betrachtet werden kann und einige

Forscher beide Terme aufgrund der schweren Abgrenzbarkeit sogar gleichsetzen (Gottfredson & Saklofske, 2009), kann die Gültigkeit der dargelegten Befunde für die vorliegende Arbeit als gegeben betrachtet werden.

Im folgenden Abschnitt wird auf die Entwicklung und Stabilität der kognitiven Fähigkeiten eingegangen, ein Aspekt, der für die vorliegende Studie, in welcher eine Kinderstichprobe untersucht wurde, von besonderer Relevanz ist.

1.2.5 Entwicklung und Stabilität

Es ist bekannt, dass die allgemeine kognitive Fähigkeit in der Bevölkerung annähernd normalverteilt ist, wobei der überwiegende Teil der Population sich um den Mittelwert von 100 IQ-Punkten clustert, während lediglich ungefähr 3% über einem Wert von 130 liegen bzw. unterhalb eines Punktwertes von 70. Doch es stellt sich die Frage nach der Entwicklung und Stabilität kognitiver Fähigkeiten im Verlauf eines Lebens. In der psychometrischen Intelligenzforschung liegt der Fokus hierbei vorrangig auf der Betrachtung quantitativer Entwicklungsschritte (Schneider, 2008). Dabei interessieren nicht nur die individuellen Zuwächse im Sinne absoluter Punktwerte. Ebenso wird der Frage nach der Stabilität der Intelligenzstruktur nachgegangen, bzw. der Stabilität von erbrachten Testleistungen. Beide Aspekte werden im folgenden Abschnitt thematisiert.

Allgemein ist ein Ansteigen der kognitiven Leistungen vom Kindesalter bis hin zum Erwachsenenalter festzustellen. Befunde aus Hirnstudien mit bildgebenden Verfahren deuten darauf hin, dass Veränderungen der kortikalen Dichte mit der Intelligenzentwicklung verknüpft sein könnten (Giedd, Schmitt & Neale, 2007).

Wie bereits in Abschnitt 1.2.3 erläutert, kann besonders gc bis ins hohe Alter Zuwächse verzeichnen, während gf ihren Höhepunkt bereits im jungen Erwachsenenalter erreicht. Dennoch kann das Konstrukt der psychometrischen Intelligenz als äußerst stabil bezeichnet werden (Gustafsson & Undheim, 1996). Dies belegen die im Zuge von Längsschnittstudien gefundenen hohen Stabilitätskoeffizienten. So fand Wilson (1986) eine Korrelation von $r = .90$ bei wiederholter Testung von Kindern bzw. Jugendlichen nach einem Jahr. Die Studie von Magnusson und Backteman (1978), in der ebenfalls Jugendliche untersucht wurden, lieferte einen Korrelationskoeffizienten von $r = .70$ bei einem Retestintervall

von fünf Jahren. Für Erwachsene ergab sich gar ein Wert von $r = .90$ bei einer wiederholten Testung im Abstand von 20 Jahren (Mortensen & Kleven, 1993). Deary, Whalley, Lemmon, Crawford und Starr (2000) berichteten einen Wert von $r = .73$ (um verminderte Streuung korrigiert) für eine 45-minütige Testung bei einem Retest-Intervall von 68 Jahren.

Die im deutschen Sprachraum durchgeführte LOGIK-Studie (Longitudinalstudie zur Genese individueller Kompetenzen; Schneider, 2008) lieferte mittelhohe Stabilitätswerte zwischen $r = .26$ und $.66$ für die sprachliche Domäne, je nach Länge des Retestintervalls. Insgesamt fallen die Stabilitätskoeffizienten höher aus, je älter die Testpersonen beim ersten Erhebungszeitpunkt und je kürzer das Retestintervall (Ahnert, Schneider & Bös, 2008).

Es ist zu beachten, dass sich Stabilität hier auf die Tatsache bezieht, dass der individuelle Rangplatz einer Testperson innerhalb einer Gruppe über die Zeit hinweg relativ konstant bleibt (Rangreihenstabilität). Dies bedeutet, dass die interindividuellen Unterschiede hinsichtlich der kognitiven Fähigkeit über die Lebensspanne betrachtet wenig Veränderung erfahren, auch wenn auf intraindividuelle Ebene durchaus Variationen auftreten, beispielsweise in Form ansteigender Rohpunktwerte (Weiß, 1998).

Für die vorliegende Arbeit ist bedeutsam, dass die Intelligenz bereits ab dem fortgeschrittenen Grundschulalter als stabil zu bezeichnen ist (Gustafsson & Undheim, 1996). Denn ab ungefähr sieben Jahren schwanken IQ-Werte lediglich in einem zu vernachlässigenden Maße (Moffitt, Caspi, Harkness & Silva, 1993). Unterhalb dieser Altersgrenze herrscht eine gewisse Variabilität der Testwerte, doch die Stabilität kognitiver Fähigkeiten ab dem mittleren Kindesalter kann als empirisch abgesichert bezeichnet werden (Bartels, Rietveld, Van Baal & Boomsma, 2002; Wilson, 1986). Die höhere Variabilität in jungen Jahren hat teilweise methodische Gründe, da in manchen Studien zu unterschiedlichen Zeitpunkten verschiedene Fähigkeiten getestet wurden. Eine weitere Ursache stellt die Tatsache dar, dass sich das genetische Potenzial immer auch vor dem Hintergrund der Umwelt entfaltet, die ein Kind umgibt, welche nun förderlich oder hinderlich sein kann.

Verhaltensgenetische Studien belegen die stark genetische Prägung kognitiver Fähigkeiten und können einen wichtigen Beitrag zum Verständnis der beachtlichen Stabilität kognitiver Fähigkeiten leisten (Hany, 2001). Zahlreiche Befunde stützen die Idee der „aktiven bzw. reaktiven Gen-Umwelt-Kovariation, wonach das Individuum

aufgrund seiner genetischen Anlagen die Umwelt zu korrespondierenden Reaktionen zwingt bzw. die Umwelt selbst verändert“ (Scarr & McCartney, 1983, zit. nach Hany, 2001, S. 82). Somit wird der Grundstein der Stabilität interindividueller Intelligenzunterschiede bereits in den ersten Lebensjahren durch die aktiv betriebene wechselseitige Beeinflussung von Genetik und Umwelt gelegt (Hany, 2001). Mit steigendem Alter wird besonders *gf* durch individuelle Faktoren beeinflusst, wie beispielsweise Krankheit. Die stärker durch formale Bildung geprägten sprachlichen Fähigkeiten können hingegen selbst bis ins höhere Alter nahezu unbeeinträchtigt bestehen bleiben (Hany, 2001). Dies erklärt die empirischen Evidenzen einer deutlicheren Rangreihenstabilität von verbalen Fähigkeiten, welche stärker mit *gc* assoziiert sind, als von räumlichen oder logischen Fähigkeiten, die eher der fluiden Intelligenz zugesprochen werden (Magnusson & Backteman, 1978; Schwartzman, Gold, Andres, Arbuckle & Chaikelson, 1987; Wilson, 1986).

Hinsichtlich der strukturellen Stabilität der allgemeinen kognitiven Fähigkeit, sprechen verschiedene Studien gegen die Annahme struktureller Differenzen zwischen unterschiedlichen Altersgruppen. Sowohl zwischen Kindern und Jugendlichen (Carroll, 1993; Kubinger & Wurst, 2000) als auch zwischen Kindern und Erwachsenen (Jäger et al., 2004) konnten keine strukturbezogenen Unterschiede nachgewiesen werden (Spengler, 2009). Allerdings ist bislang nicht abschließend geklärt, ob es Veränderungen im Hinblick auf die Relevanz einzelner Komponenten gibt (siehe Holling et al., 2004).

Zusammenfassend kann die allgemeine kognitive Fähigkeit also als äußerst stabiles Merkmal angesehen werden, auch wenn sich das Fähigkeitsniveau einer Altersgruppe über die Lebensspanne hinweg selbstverständlich ändert (Hany, 2001).

1.2.6 Zur Beziehung zwischen WM und Intelligenz

In den letzten Jahren wurde die enge Beziehung zwischen Intelligenz und WM in zahlreichen Studien untersucht. Besonders die Verbindung zwischen *gf* und WM konnte empirisch mehrfach bewiesen werden (Colom et al., 2003; Conway et al., 2002; Cowan et al., 2005; Engle, Tuholski, Laughlin & Conway, 1999; Kane et al., 2004; Oberauer, Schulze, Wilhelm & Süß, 2005; Unsworth, Redick, Heitz, Broadway, & Engle, 2009). Einige Forscher betrachten beide Konstrukte gar als isomorph

(Jensen, 1998). Besonders Studien der neunziger Jahre ließen vermuten, Intelligenz und Arbeitsgedächtnis seien identische Variablen (Kyllonen & Christal, 1990; Stauffer, Ree & Carreta, 1996). So fanden Kyllonen und Christal (1990) zwar unterscheidbare, jedoch hoch korrelierte latente Reasoning- und WM-Faktoren ($r = .80 - .90$), was sie schlussfolgern ließ: „reasoning ability is (little more than) working-memory capacity“ (S. 389).

Doch zwischenzeitlich häufen sich Belege für die weniger extreme Sicht, dass WM und Intelligenz zwar hoch korreliert, aber nicht identisch sind. So fanden Engle et al. (1999) beispielsweise lediglich Korrelationen in einem Bereich von $r = .24$ bis $.29$ zwischen WM-Tasks und Cattell's CFT (Cattell, 1973). Allerdings ist zu beachten, dass es sich hierbei um manifeste Korrelationen auf Subtestebene handelte. Eine Erweiterung dieser Studie durch Conway et al. (2002) lieferte ebenfalls moderate Korrelationen zwischen einzelnen Arbeitsgedächtnisaufgaben und dem CFT ($r = .28$ bis $.37$). Eine latente Schätzung eines WM-Faktors basierend auf drei komplexen Spannaufgaben ergab jedoch eine Korrelation von $r = .60$ mit einer latenten gf-Variable, gebildet aus dem CFT und Raven's Matrizentest, während ein latenter Faktor, basierend auf einfachen Spannaufgaben, keinen signifikanten Zusammenhang zu gf aufweisen konnte (Engle et al., 1999). Besonders erwähnenswert an dieser Studie ist, dass ein Residualfaktor, welcher um die von Arbeits- und Kurzzeitgedächtnis geteilte Varianz bereinigt wurde, ebenfalls signifikant mit gf korrelierte ($r = .49$) (siehe auch Conway et al., 2002). Diese Befunde wurden als Beweis interpretiert, dass zwischen WM und gf ein deutlich stärkerer Zusammenhang besteht, als zwischen KZG und gf. Neuere Metaanalysen stützen diese Sichtweise, da hier latente Korrelationen von $r = .50$ (Ackerman et al., 2005) bis $r = .85$ (Oberauer et al., 2005) zwischen WM und fluider bzw. allgemeiner Intelligenz berichtet wurden. Auch die aktuelle Studie von Unsworth (2010) lieferte hierzu konsistente Befunde mit einer latenten Korrelation von $r = .53$ zwischen WM und gf. Die Korrelationskoeffizienten der Beziehung zwischen KZG und Intelligenz erreichen im Gegensatz dazu nur selten Werte größer $r = .50$ (Jarrold & Towse, 2006).

Süß et al. (2002) wiesen jedoch darauf hin, dass sich in vielen Studien WM- und Reasoning-Aufgaben zu stark ähnelten, was zu einer systematischen Überschätzung der Zusammenhänge geführt haben könnte. Daher wurden Aufgaben entwickelt, die die Konstrukt-Überlappung möglichst gering halten sollten. Ein latenter WM-Faktor, der auf Basis der so konstruierten Storage-and-Processing-Aufgaben und Koordinationsaufgaben geschätzt wurde, korrelierte zu $r = .70$ mit einem aus einer Intelligenztestbatterie extrahierten Reasoning-Faktor. Auch ein extrahierter g-Faktor

wies hohe Korrelationen ($r = .65$) zu den WM-Aufgaben auf (Süß et al. 2002), so dass sich die Zusammenhänge insgesamt als stabil erwiesen.

Fasst man nun die aktuellen Befunde zusammen, wird ersichtlich, dass trotz eines substantiellen Zusammenhangs zwischen WM und Intelligenz nicht von Identität gesprochen werden kann. So stellen Conway, Kane und Engle (2003) fest, dass Arbeitsgedächtnis zwischen 30% und 50% der Varianz der Intelligenz aufzuklären vermag. Süß et al. (2002) kommentieren die prädiktive Kraft des Arbeitsgedächtnisses im Sinne eines Prädiktors der Intelligenz wie folgt: „at present, working memory capacity is the best predictor for intelligence that has yet been derived from theories and research on human cognition“ (S. 284). Laut Oberauer et al. (2005) ist die Tatsache, dass sich WM und Intelligenz als nicht identisch erwiesen haben, nur die logische Konsequenz der unterschiedlichen Ursprünge beider Faktoren, da es sich bei der psychometrischen Intelligenz um ein eher implizit hergeleitetes Konstrukt handelt, während das Konzept WM deduktiv aus theoretischen Überlegungen zur kognitiven Architektur des Menschen abgeleitet wurde. So wäre eine perfekte Korrelation schon aus theoretischer Sicht nicht zu erwarten. Dennoch ist die Untersuchung des Zusammenspiels zwischen WM und Intelligenz wertvoll, wenn es darum geht zu verstehen, was Intelligenz ist. Denn WM stellt derzeit diejenige Variable dar, die am deutlichsten mit g_f oder g korreliert, so dass die weitere Untersuchung der Arbeitsgedächtniskapazität und der Beziehung zu Intelligenz als „psychology’s best hope to date to understand intelligence“ (Oberauer et al., 2005, S.64) bezeichnet werden kann.

Obwohl die Beziehung zwischen WM und Intelligenz im Kindesalter bisher weniger intensiv beforscht wurde als bei Erwachsenen (Fry & Hale, 2000), bestätigten die verfügbaren Studien auch für diese Altersgruppe den Befund, dass es sich um zwei hoch korrelierte aber dennoch distinkte Variablen handelt (Alloway et al., 2004; Engel de Abreu et al., 2010; Fry & Hale, 2000).

So fanden De Jong und Das-Smaal (1995) eine latente Korrelation von $r = .66$ zwischen WM und g_f bei Kindern der vierten Klassenstufe. De Jong und de Jong (1996) konnten zeigen, dass die Zusammenhänge zwischen WM, Reasoning und Lesefähigkeit eine recht deutliche Stabilität in den Klassenstufen vier bis sechs aufwiesen. Auch die auf drei Jahre angelegte Längsschnittstudie von Engel de Abreu et al. (2010) ergab latente Korrelationen zwischen $r = .55$ und $.62$ zwischen einem g_f - und einem WM-Faktor in einer Stichprobe von Kindergartenkindern bzw. Erst- und

Zweitklässlern. So scheinen die Zusammenhänge beider Variablen im Kindes- und Erwachsenenalter ähnlich deutlich auszufallen.

Allerdings ist bislang sowohl bei Kindern als auch bei Erwachsenen strittig, welche Prozesse für die starke Beziehung beider Konstrukte verantwortlich sind. Während Unsworth und Engle (2005) davon ausgehen, dass exekutive Funktionen (z.B. Aufmerksamkeitskontrolle) der Korrelation zwischen gf und WM zugrunde liegen, vertreten Colom, Abad, Quiroga, Shih und Flores-Mendoza (2008) die Auffassung, dass die Beziehung zwischen WM und gf auf Kurzzeitgedächtnisprozessen bzw. der Speicherkomponente des Systems basiert. Eine aktuelle Studie von Hornung, Brunner, Reuter und Martin (2011), in der luxemburgische Vorschulkinder untersucht wurden, spricht für die von Colom und Kollegen favorisierte Sichtweise, da die Beziehung zu gf vorrangig auf die Kapazität des Kurzzeitspeichers zurückzuführen war. Allerdings wurden in dieser Untersuchung weder Aufgaben eingesetzt, die ausschließlich exekutive Funktionen beanspruchen, noch kamen komplexe visuell-räumliche Spannaufgaben zum Einsatz, was zu einer Unterschätzung des Einflusses exekutiver Prozesse geführt haben könnte (Hornung et al., 2011). Konsistent zu Engle et al. (1999) ergab jedoch die bereits erwähnte Studie von Swanson (2008), dass auch ein um KZG bereinigter WM-Faktor gf vorhersagen konnte, nicht aber KZG selbst. Auch die Studien von Bayliss, Jarrold, Baddeley, Gunn und Leigh (2005) bzw. Engel de Abreu et al. (2010) lieferten Befunde, die die Relevanz exekutiver Aufmerksamkeitskontrollfunktionen stützen. Die Ergebnisse von Tillman, Nyberg und Bohlin (2008), welche Sechs- bis Dreizehnjährige untersuchten, ergaben, dass sowohl die exekutive Kontrollkomponente als auch spezifische Kurzzeitgedächtnisprozesse Varianz in gf aufklärten. Die Autoren kamen zu dem Schluss, dass sowohl Speicher- als auch exekutive Prozesse für die Verbindung zwischen WM und Intelligenz verantwortlich sind (Tillman et al., 2008).

Auch die in Abschnitt 1.1.4 thematisierten Kernaspekte exekutiver Funktionen (Shifting, Updating, Inhibition) wurden bereits bezüglich ihrer korrelativen Zusammenhänge zu gf näher untersucht, um die Beziehung zwischen WM und gf zu beleuchten. Dabei zeigten sich Unterschiede hinsichtlich der Stärke des Zusammenhangs einzelner Komponenten zu gf. So zeigten Friedman et al. (2006), dass bei gesunden Erwachsenen Updating stärker mit Intelligenz assoziiert war als Inhibition und Shifting bzw. dass die Beziehung von Inhibition und Shifting mit Intelligenz durch die mit Updating geteilte Varianz mediiert wurde. Allerdings ist zu beachten, dass sich die Korrelationen im Kindesalter vermutlich weniger

unterscheiden, da sich, wie in Abschnitt 1.1.4 erwähnt, eine strikte Abgrenzung der drei Komponenten in jungen Jahren als schwierig erweist (Bull, Espy & Wiebe, 2008).

So kann die Frage, welche Gedächtniskomponenten und -prozesse der Beziehung zwischen WM und gf zugrunde liegen, bisher nicht abschließend beantwortet werden. Dass ein substantieller Zusammenhang besteht, gilt jedoch als unbestritten. Dabei fallen die Werte des Zusammenhangs höher aus, wenn auf latenter Ebene geschätzt wird ($r > .50$). Die dennoch auftretenden Schwankungen hinsichtlich des Korrelationskoeffizienten sind vermutlich besonders auf die von Studie zu Studie variierende Wahl der Arbeitsgedächtnisaufgaben und Intelligenztests zurückzuführen (je ähnlicher die Materialien umso höher der Wert), sowie auf die untersuchten Stichproben, welche nicht immer eine vergleichbare Fähigkeitsheterogenität aufweisen.

1.3 Motivation

Seit jeher beschäftigt sich die Wissenschaft mit der Frage, wodurch menschliches Verhalten ausgelöst und aufrecht erhalten wird (Heckhausen, 2006). Dabei gilt Motivation als eine der wichtigsten verursachenden Größen. Diese spielt auch im schulischen Kontext eine wichtige Rolle, denn es wird allgemein angenommen, dass die Leistung eines Schülers aus dem Zusammenspiel von Fähigkeit bzw. Begabung und der Motivation resultiert, sich mit dem Schulstoff zu beschäftigen. So zählt nach Schlag (2009) die Förderung der Lern- und Leistungsmotivation neben der Wissensvermittlung zu den zentralsten Aufträgen der modernen Pädagogik. Die Auseinandersetzung mit der Bedeutung der Motivation im Bildungskontext wird jedoch durch die Unübersichtlichkeit des Themengebiets erschwert. Die Vielzahl an theoretischen Konstrukten und Modellvorstellungen, die teils überlappen und sich dennoch in ihrer Operationalisierung erheblich unterscheiden, verkompliziert die Vergleichbarkeit von Studienergebnissen immens. Selbst der übergeordnete Begriff der Motivation ist nicht eindeutig definiert. Während sie im Alltagsverständnis eine homogene Instanz darzustellen scheint, von der je nach Gegebenheit mehr oder weniger zur Verfügung steht, verwendet die Wissenschaft den Begriff in einem anderen Sinne (Vollmeyer, 2005). So versteht die Motivationspsychologie unter Motivation eine nicht beobachtbare hypothetische Instanz, welche auf die Erklärung der „Zielgerichtetheit des menschlichen Handelns“ (Vollmeyer, 2005, S. 9) abzielt und zwar bezüglich „Richtung, Persistenz und Intensität“ (Heckhausen & Heckhausen, 2006, S.3). Motivation umfasst somit die Gesamtheit der in einer gegebenen Situation wirksamen Motive, welche wiederum als Beweggründe des Handelns definiert werden können und deren Befriedigung das Ziel der Handlung darstellt (Schlag, 2009). Rheinberg (2004a) definiert Motivation als die „aktivierende Ausrichtung des momentanen Lebensvollzugs auf einen positiv bewerteten Zielzustand“ (S.15) bzw. die Ausrichtung weg von negativ bewerteten Zuständen. Selbstverständlich ergibt sich die jeweilige Motivationslage nicht nur aus dem Zusammenspiel verschiedener personenbezogener Faktoren. Situative und Umweltfaktoren nehmen ebenfalls eine bedeutende Stellung ein (Heckhausen & Heckhausen, 2006). Die Vielschichtigkeit des Konstrukts kann jedoch bereits durch die Betrachtung der Bandbreite an relevanten individuellen Komponenten veranschaulicht werden, die in der Literatur Beachtung finden. Nach Rheinberg (2004b, zitiert nach Vollmeyer, 2005) spielen besonders Erwartungen, Werte, Selbstbild, Affekt und neurohormonelle Prozesse eine Rolle. Jedoch erschwert die

nicht einheitliche Verwendung dieser Begrifflichkeiten innerhalb verschiedener theoretischer Ansätze den Vergleich von Studien, die unterschiedliche Facetten beleuchten (Helmke, 1993).

Für die vorliegende Studie ist besonders der Aspekt der *Lern- und Leistungsmotivation* relevant, bei welcher die eigene Kompetenz im Zentrum der Handlung steht. Verhalten gilt dann als leistungsmotiviert, „wenn es auf die Selbstbewertung eigener Tüchtigkeit zielt, und zwar in Auseinandersetzungen mit einem Gütemaßstab, den es zu erreichen oder zu übertreffen gilt“ (Rheinberg, 2004a, S.60). Übertragen auf den Schulkontext umfasst der Begriff Leistungsmotivation all jene Handlungen, welche dem Ziel dienen, erfolgreich in der Schule zu sein (Murphy & Alexander, 2006).

Da es aufgrund der Fülle an Konzeptionen nicht möglich ist, das gesamte Spektrum der Lern- und Leistungsmotivationsforschung in dieser Arbeit wiederzugeben, erfolgt im Anschluss ein Anriss der besonders im Kontext der Bildungsforschung einflussreichsten Modelle (Kapitel 1.3.1). Dabei wird zunächst das klassische Modell von Murray (1938) angesprochen, welches als Basis vieler moderner Theorien zu verstehen ist. Daran anschließend werden die Zielorientierungstheorien, die Selbstbestimmungstheorie nach Deci und Ryan (1985) sowie die Erwartungs-Wert-Theorien thematisiert. Da diese Arbeit sich besonders auf Konstrukte des Erwartungs-Wert-Modells von Eccles und Wigfield (2002) beruft, wird dieses Modell schließlich einer detaillierteren Betrachtung unterzogen. Im Anschluss an die Modellbeschreibung erfolgt die Darstellung der Entwicklung von Fähigkeitsselbstkonzept und intrinsischer Motivation als Kernkonzepte dieser Studie (Kapitel 1.3.2). Das Kapitel schließt mit einer zusammenfassenden Betrachtung (Kapitel 1.3.3). Die Zusammenhänge zwischen Fähigkeitsselbstkonzept, intrinsischer Motivation und Schulerfolg werden in Kapitel 1.5.2 thematisiert.

1.3.1 Konzeptionen

1.3.1.1 *Need-achievement-Theorie nach Murray (1938)*

Die *need-achievement-Theorie* geht zurück auf Murray (1938), welcher sich der persönlichkeits-theoretischen Motivationsforschung verschrieb. Für ihn stellte das Bedürfnis nach Leistung ein grundlegendes menschliches Bedürfnis (*need*) dar. Der Bezug zur Persönlichkeitstheorie spiegelt sich in seinem Verständnis eines Bedürfnisses als „a more or less consistent trait of personality“ wider (Murray, 1938, S.61), welches in der Lage ist, menschliches Verhalten zu beeinflussen (Steinmayr & Spinath, 2009). Murray definierte das Bedürfnis nach Leistung als Bestreben der Überwindung von Widerständen, dem Wunsch nach Machtausübung und Bewältigung schwieriger Aufgaben (Murray, 1938). Murrays Ansatz wurde von McClelland, Atkinson, Clark und Lowell (1953) weiterentwickelt, welche die eigentliche Leistungsmotivationsforschung begründeten (Elliot & Dweck, 2005). Kern ihres Ansatzes ist die Idee, dass das Bedürfnis nach Leistung durch zwei konfligierende Einflüsse verursacht wird, nämlich der *Hoffnung auf Erfolg* und dem Wunsch, Misserfolge zu vermeiden (*Furcht vor Misserfolg*). Während Hoffnung auf Erfolg eher mit positiven Affekten wie Stolz und Zuversicht einhergeht, ist die Furcht vor Misserfolg stärker mit negativen emotionalen Zuständen assoziiert, wie beispielsweise Scham. Leistungsbezogenes Verhalten manifestiert sich dieser Theorie folgend im Spannungsfeld dieser gegensätzlichen Motive, welche sowohl die Qualität als auch die Richtung des Verhaltens beeinflussen. So aktivieren Leistungssituationen stets beide Motive, wodurch ein *Annäherungs-Vermeidungs-Konflikt* entsteht, welcher unterschiedliche Handlungsmuster bedingt, je nachdem welches Motiv dominiert (Dweck & Elliott, 1983). Sowohl Murray als auch McClelland gingen von der Generalität des Leistungsmotivs aus, weshalb sie keine domänen-spezifische Operationalisierung vornahmen. Unterschiede zwischen beiden Konzeptionen zeigen sich in McClellands stärkerer Betonung der affektiven Komponente, sowie in der Einbindung der Furcht vor Misserfolg, Aspekte die in Murrays Ansatz noch fehlen (Steinmayr & Spinath, 2009).

In Anbetracht der Häufung von Leistungssituationen im schulischen Alltag ist von einer vermehrten Aktivierung des Leistungsmotivs im Schulkontext auszugehen. So finden sich auch signifikante Korrelationen niedriger bis mittlerer Ausprägung

zwischen selbsteingeschätzten Leistungsmotiven und Schulerfolg (Paunonen & Ashton, 2001; Steinmayr & Spinath, 2009).

1.3.1.2 Zielorientierungstheorien

Die Unterscheidung eines Annäherungs- und eines Vermeidungsmotivs findet sich auch in Theorien der *Zielorientierung* wieder, welche ebenfalls im Bildungskontext stark beforscht werden. Vertreter dieser Theorien gehen davon aus, dass Leistungsverhalten auf dem bewussten Ziel basiert, die eigene Fähigkeit zu steigern bzw. diese vor sich und anderen unter Beweis zu stellen (Nicholls, 1984). Lernmotivation wird also als ein bewusstes Streben nach Lernzuwächsen verstanden (Dweck & Leggett, 1988).

Die ursprünglich dichotome Aufteilung unterschied zwischen *Aufgaben-* und *Ichorientierung* bei Nicholls (1984), bzw. zwischen *Lern-* und *Leistungszielen* bei Dweck (1986). Nicholls ging davon aus, dass sich unterschiedliche Zielorientierungen als Folge verschiedener Bewertungsmaßstäbe bezüglich eigener Fähigkeiten ergeben. Während ein intraindividueller Fähigkeitsmaßstab mit einer Aufgabenorientierung korrespondiert, führt eine interindividuelle (soziale) Bezugsnorm zu einer Ichorientierung. Dabei ist Ichorientierung darauf ausgerichtet, negative Bewertungen zu vermeiden und positives Feedback bezüglich der eigenen Kompetenz zu fördern, während aufgabenorientierte Personen die eigene Kompetenz tatsächlich zu steigern beabsichtigen, indem sie sich Aufgaben und Problemen aktiv stellen, um diese zu meistern. Dies führt zu Unterschieden im Verständnis von Erfolg. Während Ichorientierte es als Erfolg erleben, andere zu übertreffen, bedeutet Erfolg für Aufgabenorientierte die Steigerung der eigenen Kompetenz (Nicholls, 1989). Hieraus ergab sich die Vorstellung, dass sich besonders Aufgabenorientierung förderlich auf die Entwicklung intrinsischer Motivation auswirken solle. Der postulierte positive Zusammenhang fand zwar empirische Bestätigung, dennoch sind Aufgabenorientierung und intrinsische Motivation nicht gleichzusetzen (Schiefele, 2009).

Im Gegensatz zu Nicholls ging Dweck nicht davon aus, dass Zielorientierungen vorrangig aus dem differenzierten Erfolgserleben resultieren. Für ihn spielten implizite Intelligenztheorien eine viel entscheidendere Rolle. So konnte er zeigen, dass

Lernzielorientierte eher von veränderbarer Intelligenz ausgehen, welche somit Steigerungspotenzial aufweist, während Leistungszielorientierte Intelligenz eher als unveränderbares Konstrukt betrachten (Dweck, 1986).

Die dichotome Konzeption der Zielorientierung wurde schließlich durch Elliot und McGregor (2001) weiter differenziert, welche zusätzlich eine Annäherungs- und Vermeidungskomponente einbrachten. Damit kann eine Analogie zur bereits angesprochenen Unterscheidung des Erfolgs- bzw. Misserfolgsmotivs festgestellt werden (Schiefele, 2009). Elliot und McGregor (2001) unterschieden somit zwischen *Annäherungs-Lernzielen*, *Annäherungs-Leistungszielen*, *Vermeidungs-Lernzielen* und *Vermeidungs-Leistungszielen*. Während Personen mit Annäherungslernzielen auf das Erwerben von möglichst viel neuem Wissen und der Erweiterung von Kompetenzen abzielen, geht es bei Annäherungs-Leistungszielen um die Demonstration der eigenen Kompetenz gegenüber anderen. Vermeidungs-Lernziele fokussieren auf das Aufrechterhalten erworbener Fähigkeiten und Kompetenzen, Vermeidungs-Leistungsziele jedoch auf die Vermeidung der Demonstration von Inkompetenz. Die Unterscheidung dieser vier Zielorientierungen hat sich empirisch bestätigt (Elliot & McGregor, 2001). Während anfangs davon ausgegangen wurde, dass lediglich die Lernzielorientierung positiv auf Schulleistung einwirkt und Leistungszielorientierung als negativer Einfluss angesehen wurde, mehren sich die Befunde, dass die negativen Aspekte besonders der Vermeidungs-Leistungszielorientierung zuzuschreiben sind (Elliot & Church, 1997; Zusho, Pintrich & Cortina, 2005), während Annäherungs-Leistungsziele durchaus adaptiv sein können (Elliot & McGregor, 1999; Urdan, 2004).

Neuerdings zeichnet sich die Tendenz ab, eine weitere Komponente in bestehende Modelle zu integrieren. Diese als *Arbeitsvermeidung* bezeichnete Orientierung dient dem Ziel, eine Aufgabe mit so wenig Einsatz wie möglich zu bewältigen und kann somit als gegensätzlicher Pol einer hohen Leistungsmotivation aufgefasst werden (Steinmayr & Spinath, 2009). In Bezug zu Leistungskriterien zeigt sich konsistent eine negative Assoziation (Dupeyrat & Marine, 2005; Spinath, Stiensmeier-Pelster, Schöne & Dickhäuser, 2002).

Zusammengefasst können sich unterschiedliche Zielorientierungen auf Leistungsverhalten auswirken und sind damit auch im Schulkontext von Relevanz. Im Folgenden werden die empirischen Zusammenhänge jedoch nicht weiter beleuchtet, da die Skalen nicht Schwerpunkt der vorliegenden Arbeit sind.

1.3.1.3 ***Selbstbestimmungstheorie nach Deci und Ryan (1985)***

Die *Selbstbestimmungstheorie* (self-determination theory, SDT) nach Deci und Ryan (1985) stellt eine der jüngeren Konzeptionen innerhalb der Motivationsforschung dar. Kern der SDT ist die Annahme, dass Menschen bestrebt sind, einen optimalen Level an Stimulation aufrechtzuerhalten (Hebb, 1955) und dass es ein grundlegendes menschliches Kompetenzbedürfnis bzw. ein Bedürfnis nach Selbstbestimmung gibt (deCharms, 1968; White, 1959). Die SDT postuliert nun, dass Menschen aufgrund dieser Gegebenheiten stets auf der Suche nach Herausforderungen sind, da durch die Ausübung herausfordernder Aktivitäten das Kompetenzbedürfnis befriedigt werden kann. Dies gilt als Voraussetzung für die Entwicklung von intrinsischer Motivation. Intrinsische Motivation kann jedoch nur durch das Erleben von Selbstbestimmtheit aufrechterhalten werden.

Eine wichtige Neuerung der SDT gegenüber älteren Motivationstheorien ist die Unterscheidung verschiedener Arten von Motivation. So ist nach Deci und Ryan (2008) weniger das absolute Ausmaß an Motivation entscheidend für die Vorhersage wichtiger Kriterien wie Lernerfolg oder psychischer Gesundheit. Vielmehr sollte die Qualität der Motivation im Zentrum der Aufmerksamkeit stehen. Die Autoren unterscheiden *autonome* von *kontrollierter* Motivation. Je nachdem, wie autonom oder selbstbestimmt eine Handlung erlebt wird, ergibt sich entweder der Zustand der intrinsischen Motivation oder eine der vier Formen *extrinsischer Motivation* (*external*, *introjiert*, *identifiziert* und *integriert*) (Ryan & Deci, 2000). Damit werden extrinsische und intrinsische Motivation nicht mehr als antagonistische Formen verstanden, sondern als Endpole entlang eines gemeinsamen Kontinuums. Denn laut Ryan und Deci (2000) kann auch Verhalten, das vorrangig extrinsisch motiviert ist, selbstbestimmt erfolgen. Die Gemeinsamkeit von identifizierter, integrierter und intrinsischer Motivation, welche unter dem Begriff der autonomen Motivation subsumiert werden, ist nun darin zu sehen, dass sich Personen mit dem Wert einer Aufgabe identifizieren bzw. im Idealfall (intrinsische Motivation) diesen Wert in ihr Selbst integrieren, wodurch sich Gefühle der Selbstbestätigung und –bekräftigung einstellen. Kontrollierte Motivation liegt jedoch dann vor, wenn Verhalten durch externale oder introjierte Regulationsprozesse gesteuert wird. In diesem Fall wird Verhalten als Funktion von externalen Kontingenzen wie Bestrafung oder Belohnung oder Vermeidung von Scham verstanden. Im Erleben spiegelt sich dies als Gefühl wider, in einer bestimmten Weise handeln, denken oder fühlen zu müssen. Sowohl

autonome als auch kontrollierte Motivation bestimmen also menschliches Verhalten und stehen damit im Kontrast zu dem Zustand der *Amotivation*, in welchem jegliche Form von Motivation oder Intension fehlt (Deci & Ryan, 2008).

Die SDT wird derzeit bezüglich ihrer Relevanz für wichtige Themenbereiche wie Erziehungsverhalten (Joussemet, Landry & Koestner, 2008), Bildung (Guay, Ratelle & Chanal, 2008) oder Gesundheit (Miquelon & Vallerand, 2008) diskutiert, so dass zu erwarten ist, dass sie in den nächsten Jahren weitere wichtige Befunde innerhalb der Leistungsmotivationsforschung bzw. der empirischen Bildungsforschung beisteuern kann.

1.3.1.4 Erwartungs-Wert-Theorien

Eine Forschungsperspektive mit langer Tradition in der empirischen Bildungsforschung ist die *Erwartungs-Wert-Theorie*. Erwartungs-Wert-Theorien nehmen an, dass die aktuelle Handlungsmotivation durch das Zusammenspiel momentan wirksamer *Erwartungen* und *Werte* direkt bestimmt wird (Schiefele & Köller, 2006). Diesem Ansatz folgend stellen sich Personen, die überzeugt sind, eine Handlung oder Aufgabe erfolgreich bewältigen zu können, mit größerer Wahrscheinlichkeit neuen Herausforderungen, gehen neuen Problemstellungen ausdauernder nach und zeigen bessere Leistungen als Personen mit weniger günstigen Kompetenzerwartungen (Eccles & Wigfield, 2002; Jacobs, Lanza, Osgood, Eccles & Wigfield, 2002). So lässt sich die Wahl, Performanz und Persistenz einer Person bezüglich einer Handlung einerseits auf die Überzeugung zurückführen, wie gut diese Aktivität ausgeführt werden wird und andererseits auf den der Aktivität beigemessenen Wert (Atkinson, 1957; Eccles(Parsons) et al., 1983; Wigfield & Eccles, 1992). Dieser Ansatz spielt im schulischen Kontext eine wichtige Rolle, denn besonders in schwierigen Lernsituationen kann ein ungünstiges Selbstkonzept Vermeidungsverhalten auslösen, wodurch sich der Schüler vor vermeintlichen Misserfolgserlebnissen zu schützen versucht (Helmke & Weinert, 1997). Allgemein werden individuelle Unterschiede in Erfolgserwartungen und Werten als zwei der wichtigsten Prädiktoren für individuelle Differenzen in schulischen Leistungen verstanden (Eccles, O'Neill & Wigfield, 2005).

Der Aspekt subjektiver Erwartungen und Fähigkeitsüberzeugungen wird in zahlreichen Modellen und Theorien thematisiert. So postuliert die *Attributionstheorie* (Weiner, 1985), dass Personen Fähigkeit überwiegend als relativ stabiles Merkmal ansehen, das nur wenig kontrollierbar ist. Attributionen bezüglich der eigenen Fähigkeit haben nun Konsequenzen für die Motivation. Denn wird Erfolg der eigenen Fähigkeit zugesprochen, hat dies positive motivationale Effekte, während negative Konsequenzen zu erwarten sind, wenn Misserfolge auf mangelnde Fähigkeit zurückgeführt werden. Hier ist zu beachten, dass Nicholls (1978) entwicklungsbedingte Veränderungen hinsichtlich der Attribution von Fähigkeit annimmt (siehe Abschnitt 1.3.2). Als weiterer Ansatz wäre Covingtons (1992) *Selbstwerttheorie* zu nennen, nach der Menschen bestrebt sind, eine positive Fähigkeitsüberzeugung aufrecht zu erhalten, um dadurch den Selbstwert zu schützen. Auch die in Abschnitt 1.3.1.3 angesprochene Selbstbestimmungstheorie nach Deci und Ryan (1985) oder die *Selbstkonzeptforschung* (Marsh, 1990) betrachten individuelle Fähigkeitsüberzeugungen als zentrale Aspekte menschlicher Motivation. Doch jeder Ansatz bezieht sich auf eine eigene Terminologie. So finden sich in der Literatur Begriffe wie *leistungsbezogenes Selbstkonzept*, *Erfolgs- und Misserfolgserwartung*, *Fähigkeitsselbstbild*, *Selbstwirksamkeit* oder *Selbstvertrauen* (Eccles & Wigfield, 2002). Auch wenn allen Variablen eine Einschätzung subjektiver Fähigkeiten gemein ist und sie auf einen gemeinsamen Ursprung zurückzuführen sind, findet dennoch in der Literatur eine getrennte Betrachtung dieser Konstrukte statt, da sie mit unterschiedlichen Spezifizierungsniveaus und Operationalisierungen einhergehen. So variieren auch die Zusammenhänge der Variablen mit leistungsbezogenen Kriterien, beispielsweise weil Variablen, die domänen-spezifisch betrachtet werden, üblicherweise höhere Zusammenhänge aufweisen als domänen-unspezifische (Valentine, DuBois & Cooper, 2004).

Die soeben beschriebenen subjektiven Erwartungen stellen nur eine auslösende Teilkomponente dar, die Menschen dazu veranlasst, eine Aktivität auszuführen. Es ist jedoch nicht hinreichend, dass eine Person die eigene Fähigkeit positiv einschätzt, die zur Bewältigung einer Aufgabe notwendig ist. Der empfundene Wert entscheidet maßgeblich mit, ob einer Aktivität nachgegangen wird oder nicht (vgl. Eccles & Wigfield, 2002, S. 112). Besonders das bereits angesprochene Konzept der *intrinsischen Motivation*, sowie der Faktor *Interesse* als „mit subjektiven Werten verbundene Präferenz für bestimmte Lerngegenstände“ (Wild & Krapp, 1995, S.591), nehmen in der Literatur einen großen Stellenwert ein. Gängigerweise wird intrinsische Motivation als Absicht definiert, eine Handlung „um ihrer selbst willen

durchzuführen, weil diese z. B. als interessant, spannend, herausfordernd usw. erscheint“ (Schiefele & Schreyer, 1994, S. 1 - 2). Es ist keine Verstärkung von außen erforderlich, denn die Handlung selbst wird als belohnend empfunden, wodurch sich Freude und Vergnügen einstellen (Eccles & Wigfield, 2002; Schiefele & Köller, 2006). Während also intrinsische Motivation die Ausführung einer Handlung ohne Ausrichtung auf bestimmte Konsequenzen darstellt, bedeutet extrinsische Motivation, Handeln mit dem Ziel gewisse positive Konsequenzen zu erfahren bzw. negative zu vermeiden (Schiefele & Köller, 2006). Intrinsische Motivation ergibt sich also von innen heraus, bei extrinsisch motivierten Handlungen steuert ein außerhalb der Aktivität liegendes Motiv, beispielsweise eine Belohnung (Eccles & Wigfield, 2002). Die Relevanz der intrinsischen Motivation für den schulischen Kontext liegt auf der Hand, da sich diese als Lernfreude manifestiert und mit Neugier, Ausdauer und dem Wunsch, Neues zu lernen, assoziiert ist (Gottfried, 1990).

Ein Modell, welches den modernen Erwartungs-Wert-Ansätzen zuzuordnen ist und sich in der empirischen Bildungsforschung aufgrund seiner guten Anwendbarkeit auch bei jüngeren Stichproben bewährt hat, ist das Modell der Leistungsmotivation von Eccles und Kollegen (Eccles & Wigfield, 2002). Fähigkeitsüberzeugungen und Werte werden in diesem Modell domänen-spezifisch erfasst. Da die für diese Arbeit zentralen motivationalen Komponenten *Fähigkeitsselbstkonzept* (FSK) und *intrinsische Motivation* (IM) in Anlehnung an dieses Modell operationalisiert wurden, wird es im folgenden Abschnitt detaillierter beschrieben.

1.3.1.5 Das Erwartungs-Wert-Modell von Eccles und Wigfield (2002)

Das Erwartungs-Wert-Modell von Eccles und Kollegen (2002) gilt als einflussreiches und empirisch abgesichertes Modell (Eccles, 2005; Wigfield & Eccles, 2000). Das Modell integriert Erwartungs- und Wertkomponenten und beschreibt die verschiedenen Einflussfaktoren und Wirkzusammenhänge auf Performanz und leistungsrelevante Auswahlentscheidungen, wobei sich letzteres beispielsweise auf Karriereentscheidungen oder Kurswahl bezieht. Eccles und Kollegen griffen bei der Modellentwicklung bestehende Ansätze auf (beispielsweise Atkinson, 1957), allerdings erweiterten sie diese um Sozialisierungseinflüsse und zusätzliche psychologische Variablen. Abbildung 1 zeigt die aktuelle Modelldarstellung.

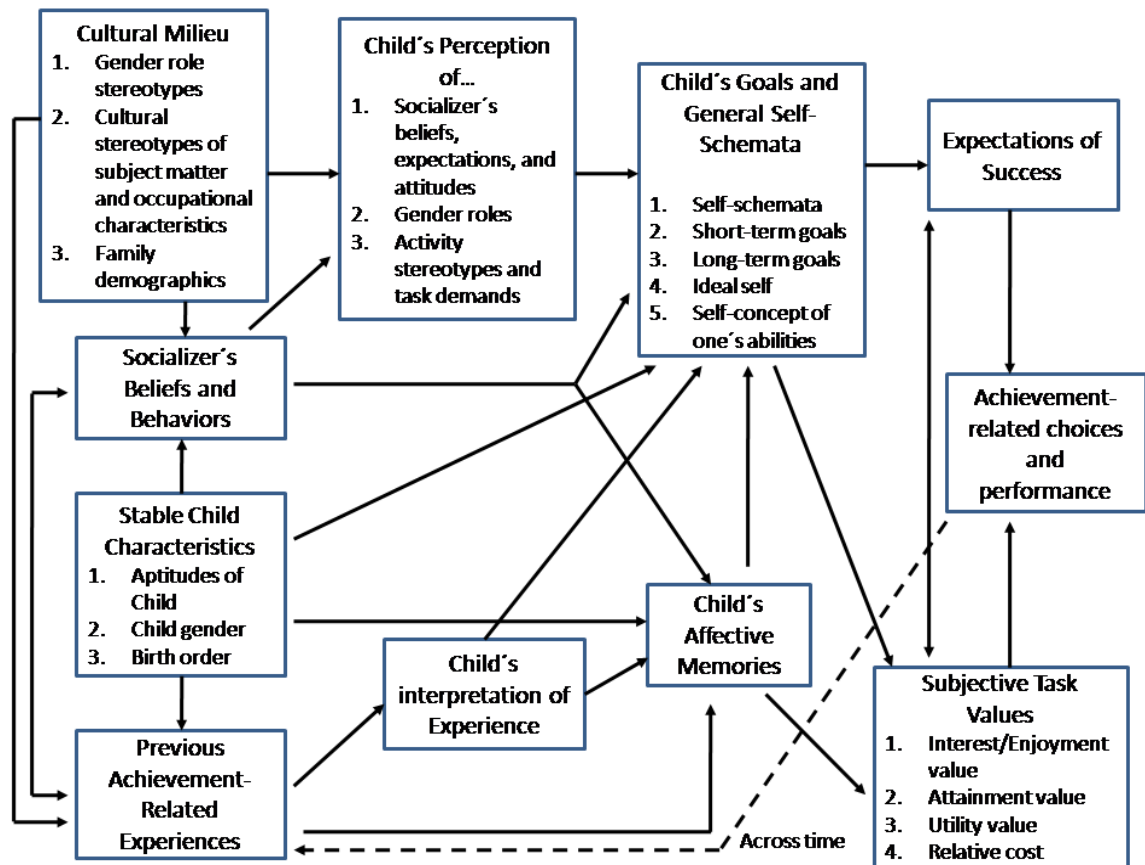


Abbildung 1: Erwartungs-Wert-Modell der Leistungsmotivation (Eccles & Wigfield, 2002)

Als Vorteile des Modells gegenüber anderen Konzeptionen sind besonders die bereichsspezifische Operationalisierung der Konstrukte, sowie das hohe Maß an Spezifität zu sehen (Wigfield & Eccles, 2002). Wie Abbildung 1 illustriert, nehmen subjektive Erfolgserwartungen und subjektiver Aufgabenwert zentrale Positionen im Modell ein, denn durch sie werden leistungsrelevante Entscheidungen, die Leistung selbst, sowie Fleiß und Persistenz direkt beeinflusst. Entscheidend ist, dass das Modell die besondere Rolle der subjektiven Einschätzung heraus stellt, die nicht zwangsläufig realistisch sein muss. Erwartungen und Werte wiederum werden durch individuelle Faktoren verändert, wie persönliche Ziele, Selbstschemata oder affektive Gedächtnisrepräsentationen (Wigfield & Eccles, 2000). Diesen sozial-kognitiven Einflussfaktoren sind individuelle Verarbeitungsprozesse vorgeschaltet, die beispielsweise durch persönliche Erfahrungen geprägt werden. Außerdem übt die persönliche Wahrnehmung der Erwartungen und Überzeugungen anderer einen Einfluss aus. Somit betont das Modell die Rolle von Sozialisationsprozessen, da das Verhalten von Bezugspersonen und die Wahrnehmung dieses Verhaltens die individuelle Bewertung einer Erfahrung und die persönlichen Ziele und

Selbstschemata direkt beeinflussen. Natürlich spielen auch kulturelle bzw. gesellschaftliche Umweltbedingungen eine Rolle (Eccles et al., 1998). Insgesamt gelten Aufgabencharakteristika (Schwierigkeit, Zeitaufwand), Ziele und Selbstwahrnehmungen als bedeutendste Wirkfaktoren auf den subjektiven Wert einer Aufgabe, während Erfolgserwartungen besonders durch das Fähigkeitsselbstkonzept und den Schwierigkeitsgrad der Aufgabe beeinflusst werden (Eccles et al., 1983).

Erwartungen werden von Eccles et al. (1983) als die Überzeugung einer Person definiert, wie gut eine anstehende Aufgabe bewältigt werden wird. Damit erhält die Dimension der Erfolgserwartung einen auf Zukünftiges ausgerichteten Fokus. Die den Erfolgserwartungen vorgelagerten Fähigkeitsüberzeugungen umfassen die Bewertung der eigenen Kompetenz in verschiedenen Gebieten. Empirisch hat sich allerdings die Trennung dieser auf zwei Ebenen befindlichen Überzeugungsformen weder für Kinder noch für Erwachsene bestätigt (Eccles & Wigfield, 1995).

Typischerweise erfolgt die Repräsentation der Erwartungskomponente in Form des fachspezifischen FSK, so auch in dieser Arbeit. Während das Selbstkonzept allgemein einer hierarchischen Struktur unterliegt, mit einem an der Spitze stehenden generellen Selbstkonzept und darunter liegenden spezifischeren Fähigkeitsselbstkonzepten (z.B. akademisches Selbstkonzept) (siehe Shavelson, Hubner & Stanton, 1976), bezieht sich das in dieser Studie operationalisierte FSK ausschließlich auf das schulische Selbstkonzept, welches als Wahrnehmung und Bewertung eines Schülers „of doing well or poorly in school in general (general academic self-concept) or in particular academic areas, e.g. mathematics self-concept“ (Skaalvik & Skaalvik, 2004, S. 1188) definiert werden kann (siehe Kapitel 3.4.2).

Auch die subjektiven Aufgabenwerte werden im Modell näher spezifiziert. Eccles und Wigfield (2002) benennen vier motivationale Wertkomponenten, nämlich *Wichtigkeit* (attainment value), *Nützlichkeit* (utility value), *intrinsischer Wert* (intrinsic value) und *Kosten* (cost). Wichtigkeit bezieht sich darauf, wie wichtig es für eine Person ist, eine bestimmte Leistung zu erbringen oder eine Aufgabe erfolgreich auszuführen (Eccles, 2005). Dieser Wert liegt besonders dann in hoher Ausprägung vor, wenn die Aufgabe es erlaubt, saliente Anteile des Selbstschemas zu bestätigen (Eccles, 1984). Die subjektive Nützlichkeit ergibt sich aus der Passung zwischen Aufgabe und aktuellen bzw. zukünftigen Zielen, wie beispielsweise Karriereplänen und umfasst damit eher externale Gründe der Handlungsausführung (Eccles et al., 1998). So kann eine

Aufgabe als nützlich im Sinne einer Zielerreichung sein, aber gleichzeitig als uninteressant empfunden werden. Dies wäre beispielsweise im Falle eines Studenten gegeben, der Seminare besucht, die ihn zwar nicht unmittelbar interessieren, ohne die er aber seinen Abschluss nicht erhält. Der intrinsische Wert einer Aufgabe hingegen meint das Vergnügen, das sich durch das bloße Ausführen einer Aufgabe beim Handelnden einstellt. Diese Wertkomponente deckt sich stark mit dem bereits angesprochenen Konzept der intrinsischen Motivation nach Deci und Ryan (1985), sowie den Konstrukten Interesse und Flow nach Csikszentmihalyi (1988) oder Schiefele (1991) (Eccles et al., 1998) und wird in der vorliegenden Studie als Repräsentant der Wertkomponente herangezogen (siehe Kapitel 3.4.2).

Da jede Entscheidung für eine bestimmte Aufgabe bzw. die Ausführung dieser auch mit negativen Aspekten verknüpft ist, wie etwa einem gewissen Mindestaufwand, Versagensangst oder der Limitierung des Zugangs zu Alternativmöglichkeiten, postulieren Eccles et al. (1998) Kosten als vierte relevante aufgabenspezifische Wertkomponente. Während die ersten drei Komponenten eher zu einer positiven Aufgabenbewertung beitragen, führt der Kostenfaktor zu einer negativen Valenz der Aufgabe (Eccles, 2005). Entscheidungen werden also durch das Verhältnis von Kosten und Nutzen stark beeinflusst (Eccles & Wigfield, 2002).

Insgesamt sprechen viele empirische Befunde für das Modell von Eccles und Kollegen, dessen Anwendbarkeit sowohl bei Grund- als auch bei Sekundarschülern nachgewiesen werden konnte (Eccles, 2005; Eccles & Wigfield, 2002; Eccles, Wigfield, Harold & Blumenfeld, 1993). So konnten FSK und Erfolgserwartungen als Prädiktoren schulischer Leistung für Mathematik und Englisch bestätigt werden (Eccles & Wigfield, 2002). Die aufgabenspezifischen Werte erwiesen sich als besonders hilfreich zur Prognose der Kurswahl in vielen Fächern, selbst bei Kontrolle der vorherigen Leistung (Eccles, Adler & Meece, 1984; Eccles et al., 1998).

Als Kritik an diesem Modell bzw. an der Erwartungs-Wert-Theorie allgemein, wird häufig angeführt, dass die Entstehung einer motivationalen Tendenz zu sehr von den antizipierten Konsequenzen einer Handlung abhängt (Schiefele & Streblow, 2005). Es wird argumentiert, dass die vorrangig rational-kognitive Betrachtung motivationaler Prozesse der menschlichen Natur nicht gerecht wird, da Erwartungen und Werte nicht ausschließlich durch logische Entscheidungsprozesse beeinflusst werden, sondern häufig einfachere, optimistische Strategien zum Einsatz kommen (Eccles et al., 1998). Allerdings ist zu beachten, dass sowohl die Erwartungs- als auch die Wertkomponenten nicht zwangsläufig in Form bewusster Repräsentationen

vorliegen müssen (Schneider & Schmalt, 2000). Außerdem wird im Modell von Eccles und Kollegen (1998) durch die Hinzunahme von affektivem Gedächtnis und kulturspezifischen Einflüssen weniger bewussten und damit weniger rationalen Entscheidungsprozessen Raum gegeben (Eccles et al., 1998).

Schließlich rechtfertigt die Fülle an empirischen Befunden zur Beziehung zwischen Erwartungen, Werten und Performanz die Weiterverfolgung des Modells (Eccles et al., 1998). Auf diese wird in Abschnitt 1.5.2 näher eingegangen. Im Anschluss erfolgt nun die Darstellung wichtiger Befunde bezüglich der Entwicklung der Erwartungs- und Wertkomponenten.

1.3.2 Entwicklung und Stabilität der Erwartungs- und Wertkomponenten

Der Grundstein der Leistungs- und Lernmotivation wird bereits im frühen Kindesalter gelegt, wobei Sozialisationsbedingungen entscheidend beeinflussen, welche Möglichkeiten der Entwicklung und Förderung des Leistungsstrebens eines Kindes realisiert werden können (Schlag, 2009). Neben der Entwicklung theoretischer Motivationsmodelle beschäftigt sich somit ein Teilgebiet der Motivationsforschung auch mit der Frage, wie sich sowohl die Leistungsmotivation allgemein als auch spezifische Konzepte wie Erwartungen, Werte und Ziele entwickeln und im Laufe des Lebens verändern, bzw. welche Mechanismen zur Entstehung individueller Unterschiede beitragen. Diese Aspekte sind besonders für die pädagogische Praxis relevant, wenn es darum geht, die Leistungsmotivation der Schüler zu fördern bzw. aufrechtzuerhalten. Daher werden im Folgenden empirische Befunde zusammengefasst, die sich auf die Entwicklung der für die vorliegende Studie relevanten Faktoren FSK und IM beziehen.

1.3.2.1 Entwicklung und Stabilität des Fähigkeitsselbstkonzepts

Die Entstehung und Entwicklung motivationaler Aspekte wie FSK und IM ist an die Fähigkeit geknüpft, eigenes Verhalten selbstreflexiv betrachten und bewerten zu können. Hinsichtlich des erstmaligen Auftretens von selbstbewertenden

Verhaltensweisen konnte Heckhausen (1984, 1987) bereits in den Achtzigern zeigen, dass schon Kinder ab zweieinhalb Jahren imstande waren, auf nonverbaler Ebene Freude nach Erfolg bzw. Enttäuschung nach Misserfolg zu zeigen. Einige Monate später folgten Expressionen von Scham und Stolz. In Wettbewerbssituationen mit anderen Kindern traten Freude und Trauer bei Drei- bis Vierjährigen auf (Eccles et al., 1998). Auch Stipek, Recchia und Mc Clintic (1992) konnten eine Verknüpfung von kindlicher Reaktion nach erfolgreicher Aufgabenbewältigung und wertschätzendem Verhalten anderer bei zweijährigen Kindern nachweisen. Ab einem Alter von drei und mehr Jahren waren die Kinder schließlich in der Lage, ihr Verhalten bzw. ihre Leistung auch ohne die Reaktion von Erwachsenen selbst zu bewerten (Eccles et al., 1998). Diese Studien zeigen, dass bereits Vorschulkinder zwischen Erfolg und Misserfolg differenzieren und entsprechend reagieren können. Die Konzepte, die sich in diesem Alter manifestieren, stellen vermutlich die Grundlage für die später ausgereiften motivationalen Eigenschaften eines Kindes dar (Eccles et al., 1998).

Hinsichtlich der faktoriellen Struktur von Fähigkeitsüberzeugungen konnte mehrfach belegt werden, dass diese ab dem Zeitpunkt der Erfassbarkeit domänen-spezifisch vorliegen, selbst im Vorschulalter (Eccles et al., 1993; Wigfield, Eccles & Pintrich, 1996). Zwar finden sich häufig positive Korrelationen zwischen den Fähigkeitseinschätzungen über Domänen hinweg (Eccles et al., 1993; Wigfield & Eccles, 1992), aber dennoch betreffen entwicklungsbedingte Veränderungen weniger die Faktorenstruktur, sondern vielmehr die Art und Weise, wie Kinder Fragen zu subjektiven Fähigkeitsüberzeugungen beantworten. Während junge Kinder besonders dazu neigen, in Extremen zu antworten, wodurch die Varianz sehr eingeschränkt wird und sich nur geringe Korrelationen zu Eltern- und Lehrerratings finden (Wigfield et al., 1996), verfeinern sich die Antworten mit steigendem Alter. So steigen die Zusammenhänge zwischen der Einschätzung der Kinder und Leistungsindikatoren im Grundschulalter deutlich an (Eccles et al., 1998). Eine weitere wichtige Erkenntnis betrifft die Mittelwertsveränderungen des FSKs. So ist bereits ab dem höheren Grundschulalter ein Absinken des FSKs festzustellen, ein negativer Trend, der auch nach dem Übergang in die weiterführenden Schulen anhält (Dweck & Elliott, 1983; Guay, Marsh & Boivin, 2003; Spinath & Spinath, 2005a). Eccles et al. (1993) belegten diese Abwärtsentwicklung bei Sieben- bis Zehnjährigen für verschiedene Fächer und Kompetenzen wie Mathematik, Musik und Lesen. Weiter Belege finden sich bei Jacobs et al. (2002), die Kinder von der ersten bis zur zwölften Klassenstufe untersuchten. Allerdings handelte es sich bei dieser Stichprobe aufgrund einer Überrepräsentation höherer Gesellschaftsschichten um eine selektive,

wodurch die Vergleichbarkeit eingeschränkt wurde. Doch auch die von Spinath und Steinmayr (2008) an Grundschulkindern durchgeführte Studie ergab konsistente Befunde.

Es stellt sich Frage nach den Ursachen dieser negativen Entwicklung. Ein Aspekt ist sicherlich, dass sich im Laufe der Schullaufbahn, beispielsweise durch die Einführung der Schulnoten, mehr soziale Vergleichsmöglichkeiten ergeben als im Kindergarten, wodurch sich die Fähigkeit der Kinder, Feedback zu verstehen und zu interpretieren, stets verbessern dürfte. Daraus resultieren schließlich realistischere und differenziertere Selbsteinschätzungen (Eccles et al., 1998, Wigfield & Eccles, 2000). In der Schulumgebung werden außerdem Bewertungen und Wettbewerbssituationen insgesamt salient, weshalb sich einige Kinder mit zunehmendem Alter schlechter einschätzen (Eccles, Midgley & Adler, 1984).

Wichtig für die Betrachtung der Entwicklungsprozesse ist weiterhin die Frage, inwiefern sich das kindliche Verständnis von Fähigkeitsüberzeugungen im Laufe der Zeit verändert. So untersuchten Nicholls et al. (1986) das Verständnis von Kindern über Fähigkeit, Intelligenz, Fleiß und Aufgabenschwierigkeit. Genauer ging es den Autoren um die Frage, wie Kinder die verschiedenen Variablen mit unterschiedlichen Leistungsniveaus in Verbindung setzen. Aus ihren Befunden schlossen Nicholls et al. (1986), dass Kinder zwischen fünf und sechs Jahren nicht klar zwischen Fleiß, Fähigkeit und Leistung im Sinne von Ursache-Wirkungs-Zusammenhängen unterscheiden. Kinder zwischen sieben und neun Jahren führen Leistung vor allem auf Fleiß und Einsatz zurück. Im Alter von neun bis zwölf Jahren beginnen Kinder, Fähigkeit und Fleiß als trennbare Ursachen zu begreifen, doch eine beständige Separierung von Fleiß und Fähigkeit und dem Verständnis von Fähigkeit als Kapazität liegt erst im jungen Erwachsenenalter vor. Hier verstehen Kinder, dass Fähigkeit als limitierende Größe anzusehen ist, die durch Fleiß nur innerhalb gewisser Grenzen zu kompensieren ist, bzw. dass das Maß an Fleiß, das zur Erbringung einer Leistung erbracht werden muss, auf das Fähigkeitsniveau schließen lässt (Kun, Parsons & Ruble, 1974). Dass besonders das verinnerlichte Konzept bezüglich Intelligenz oder Fähigkeit Auswirkungen auf motivationale Prozesse hat, konnten Nicholls (1984) und Dweck (Dweck & Elliott, 1983; Dweck & Leggett, 1988) belegen. Bei Kindern, die Intelligenz als stabile, unveränderliche Größe betrachteten (*entity view*), erhöhte sich die hemmende Wirkung von Misserfolgen auf die Motivation, da die Kinder nicht erwarteten, durch mehr Einsatz ihre Leistung verbessern zu können. Eine inkrementelle Sicht jedoch, nach der Intelligenz veränderbar und somit steigerbar ist, sollte der Entwicklung von Mustern wie erlernter

Hilflosigkeit als Reaktion auf Misserfolge entgegenwirken, da die Hoffnung besteht, durch mehr Einsatz die Leistung noch verbessern zu können (Dweck & Master, 2009). Spinath und Steinmayr (2008) weisen darauf hin, dass das zunehmende Verständnis von Intelligenz als stabile Größe ebenfalls eine Ursache des Abfalls der Fähigkeitsüberzeugungen darstellt.

Hinsichtlich der Entstehung und Veränderung des FSKs ist weiterhin zu erwähnen, dass sowohl intra-individuelle als auch soziale Vergleichsprozesse eine Rolle spielen. Das *Internal/External-frame of reference model* (Marsh, 1986) besagt, dass die Einschätzung eines Schülers bezüglich der eigenen Fähigkeit durch das Zusammenspiel zweier Vergleichsprozesse zustande kommt. Einerseits beeinflussen externale Vergleichsprozesse die jeweilige Einschätzung, da die eigene Leistung mit der Leistung von Klassenkameraden in Bezug gesetzt wird. Doch auch interne Prozesse spielen eine Rolle, da die eigenen Leistungen auch zwischen verschiedenen Fächern verglichen werden. Je nachdem, welcher Prozess stärker gewichtet wird, ergeben sich durchschnittliche, überdurchschnittliche oder unterdurchschnittliche Fähigkeitsüberzeugungen (Marsh, Kong & Hau, 2000). Wie einflussreich der jeweilige Bezugsrahmen auf die Entwicklung bzw. Veränderung des FSKs sein kann, zeigt die Forschung zum *Big-Fish-Little-Pond-Effekt* (BFLPE) (Marsh, 1987; Marsh et al., 2008). Da die eigenen Leistungen stets auch mit den Leistungen der Bezugsgruppe verglichen werden, entwickeln Schüler mit objektiv gleicher Fähigkeit ein höheres FSK, wenn die durchschnittliche Leistung der Bezugsgruppe niedriger ist als die eigene, während ein niedrigeres FSK entstehen kann, wenn die Leistung der Bezugsgruppe über der eigenen liegt (Marsh et al., 2008).

1.3.2.2 Entwicklung und Stabilität der subjektiven Wertkomponenten

Eccles und Kollegen (1998) konnten mehrfach belegen, dass Kinder bereits in der ersten Klasse innerhalb einer Domäne zwischen subjektiven Aufgabenwerten und Kompetenzüberzeugungen differenzieren. Zwar fanden sich innerhalb einer Domäne positive Korrelationen zwischen Fähigkeitsüberzeugungen und Aufgabenwerten (Eccles & Wigfield, 1995; Gottfried, 1985), aber dennoch lagen beide Variablen als distinkte Faktoren vor. Allerdings bildet sich die Differenzierung verschiedener Wertkomponenten erst im Laufe der Grundschulzeit heraus. Während in der frühen

Grundschule nur zwischen Interesse und Wichtigkeit unterschieden wurde, differenzierten Kinder der fünften bis zwölften Klassenstufe zwischen Wichtigkeit, Interesse und Nützlichkeit (Eccles et al., 1998).

Ähnlich wie für FSK findet sich auch für IM und andere Wertkomponenten ein Absinken im Verlauf der Schulzeit. Als mögliche Ursache des Absinkens nennt Baumert (1995), dass Heranwachsende sich immer neue Erfahrungsbereiche erschließen, wodurch neue Interessen generiert werden und der Einfluss der Schule abgeschwächt wird (Eder, 1992). Darüber hinaus üben zahlreiche Kontextfaktoren wie das Lehrerverhalten im Klassenraum, Unterrichtsstil oder Leistungsdruck einen Einfluss aus (Eccles & Midgley, 1989).

Hinsichtlich der Beziehung zwischen FSK und IM stellt sich die Frage, ob die Tatsache, dass beide Variablen im Laufe der Schulzeit absinken, auf einen kausalen Zusammenhang zwischen beiden Faktoren hindeutet. Wigfield (1994) stellte fest, dass bei sehr jungen Kindern Aufgabenwerte und Fähigkeitsüberzeugungen noch wenig assoziiert sind, weshalb sehr junge Kinder noch vermehrt Tätigkeiten nachgehen, die sie interessant finden, selbst wenn sie nicht gut darin sind. Erst mit zunehmendem Alter verknüpfen sich beide Dimensionen nach und nach, wodurch sich schließlich ein Gefühl der Selbstwirksamkeit manifestieren kann (Eccles, Wigfield & Blumenfeld, 1984).

Der korrelative Zusammenhang zwischen beiden Variablen konnte mehrfach empirisch nachgewiesen werden. Beispielsweise fanden Steinmayr und Spinath (2009) Korrelationen von $r = .81$ in Mathematik und von $r = .55$ in Deutsch, was die Wichtigkeit der fachspezifischen Betrachtung verdeutlicht.

Viele der im vorherigen Abschnitt dargestellten Theorien legen eine kausale Verknüpfung von FSK und IM nahe, weshalb eine Veränderung in der Ausprägung des FSK eine Änderung der IM nach sich ziehen sollte (Deci & Ryan, 1985; Wigfield & Eccles, 2000). Wäre dies zutreffend, würde das im Laufe der Schulzeit einsetzende Absinken des FSK also das Absinken der IM verursachen. Der empirische Nachweis dieses kausalen Zusammenhangs konnte jedoch bisher nicht erbracht werden. So ergaben alle längsschnittlichen Studien, die sich mit dieser Fragestellung beschäftigten, dass entweder kein Effekt gefunden werden konnte oder lediglich in einem zu vernachlässigenden Umfang (Marsh, Trautwein, Lüdtke, Köller & Baumert, 2005; Skaalvik & Valas, 1999; Spinath & Spinath, 2005; Spinath & Steinmayr, 2008). Somit zeichnet sich ab, dass trotz der korrelativen Beziehung zwischen FSK und IM eher von einer parallelen Entwicklung denn von einer kausalen Beeinflussung

auszugehen ist (Spinath & Steinmayr, 2008). Dies erklärt sich durch die Tatsache, dass bestimmte äußere Faktoren, wie beispielsweise eine schlechte Note, gleichzeitig auf beide motivationalen Facetten einwirken (Spinath & Spinath, 2005). Im Hinblick auf die praktische Relevanz dieser Befunde ist anzumerken, dass es durchaus interindividuelle Varianz hinsichtlich des Absinkens der motivationalen Komponenten gibt, sodass sich die Frage stellt, für welche Kinder das Risiko besonders hoch ist, im Laufe der Schulzeit an Motivation bzw. Fähigkeitsüberzeugung einzubüßen (Spinath & Steinmayr, 2008). Diese Fragestellung stellt eine lohnenswerte für die weitere Forschung dar.

1.3.3 Zusammenfassung

Das wissenschaftliche Verständnis des Konstrukts Motivation hat sich im Laufe des 20. Jahrhunderts dramatisch verändert. Ursprünglich vorrangig durch Triebtheorien geprägt, entwickelten sich nach und nach behaviorale und schließlich kognitiv-konstruktivistische Ansätze. Die handelnde Person wurde zunehmend als planvoller, zielgerichteter Akteur verstanden, dessen Aufgabe darin besteht, unterschiedliche Ziele je nach Kontext abzuwägen bzw. konfligierende Ziele zu vereinbaren. Neuerdings ist eine Wiederbelebung des Versuchs emotionale und unbewusste Prozesse in bestehende Konzeptionen zu integrieren festzustellen. Außerdem wird der Einschluss kontextueller Faktoren im Rahmen umfassender Netzwerk-Modelle forciert (Eccles et al., 1998).

Eine Herausforderung für die zukünftige Motivationsforschung ist sicherlich die Integration der Fülle an Forschungsbefunden, die sich teilweise auf sehr ähnliche Konstrukte beziehen, jedoch aufgrund einer Vielzahl an unterschiedlichen Operationalisierungen und einer unüberschaubaren Zahl an Messinstrumenten nur schwer in eine übergreifende Theorie einzubetten sind.

1.4 Sozioökonomischer Status

Eine Familienvariable, die immer wieder hinsichtlich ihres Einflusses auf schulische Erfolge diskutiert wird, ist der familiäre *sozioökonomische Status* (socio-economic status, SES). Wie aktuell diese Frage ist, zeigt der bereits erwähnte Artikel der GEW, welcher unter der Überschrift „Arbeiterkinder müssen Superschüler sein...um nach der Grundschule aufs Gymnasium zu gehen“ veröffentlicht wurde (Demmer, 2009). Der Titel dieses Artikels spiegelt die weit verbreitete Sorge wider, dass in unserem Schulsystem Kinder aus niedrigeren sozialen Schichten aufgrund ihrer Herkunft schlechtere Chancen haben, einen hohen Bildungsabschluss zu erlangen, als Kinder aus Familien höherer Schichten. Forciert wurde diese Befürchtung auch durch großangelegte Vergleichsstudien, wie beispielsweise die PISA-Untersuchung (OECD, 2001). Denn ein Befund dieser Untersuchung war, dass sich gerade in Deutschland die soziale Herkunft erheblich auf den Schulerfolg auswirkt (Baumert & Schümer, 2001). Aufgrund der hohen sozialpolitischen und gesellschaftlichen Relevanz der Frage nach dem Einfluss des SES, sollen im Folgenden wichtige Aspekte bezüglich Erfassung und Bedeutung des SES dargelegt werden. Die Zusammenfassung wichtiger empirischer Befunde zum Zusammenhang zwischen SES und Schulerfolg erfolgt in Kapitel 1.5.3.

Auch wenn zu beachten ist, dass in der Literatur eine Vielzahl an Variablen und Definitionen unter dem Begriff des SES aufscheint (White, 1982), herrscht doch das weit verbreitete Verständnis, dass der SES die Verortung einer Person oder Familie in der sozialen Hierarchie einer Gesellschaft indiziert (Baumert & Maaz, 2006). Allgemein gelten Einkommen, Bildung und Beruf als die drei wichtigsten Indikatoren des SES, wobei besonders häufig die elterliche Schulbildung erfasst wird (Ensminger & Fothergill, 2003). Dabei gilt üblicherweise der höchste Schulabschluss eines Elternpaares als Indikator der Bildung. Das Einkommen wird meist in Form einer Gehaltsabfrage erfasst, während schließlich mit Hilfe verschiedener Klassifikationssysteme unterschiedliche Berufe in eine Rangordnung hinsichtlich Stellung und Prestige überführt werden können (Hoffmeyer-Zlotnik, 2003). Verbreitet ist außerdem die Zusammenfassung der drei gängigen Indikatoren Bildung, Einkommen und Beruf zu einem Gesamtindex, wobei Duncan und Magnuson (2003) darauf hinweisen, dass durch die Verwendung eines solchen Index spezifische Effekte der einzelnen Komponenten übersehen werden könnten, weshalb sie eine getrennte Betrachtung anraten.

Unterschiede im SES einer Familie gehen mit unterschiedlichen finanziellen Ressourcen einher, aber auch mit Differenzen bezüglich nicht materieller Aspekte wie Einfluss und Prestige (Baumert & Maaz, 2006). Die finanziellen Möglichkeiten einer Familie beeinflussen die Ausgestaltung der Familienumwelt, sowohl bezüglich der unmittelbaren Wohnumgebung, als auch hinsichtlich der Aktivitäten und Unternehmungen, die realisiert werden können. Aufgrund der Vorbildfunktion von Eltern werden durch gemeinsame Aktivitäten, aber auch durch Bereitstellung gewisser bildungsfördernder Umwelten, die notwendigen Einstellungen und Werte vermittelt, die schließlich durch Förderung von Interesse auch die Entwicklung schulischer und außerschulischer Kompetenzen positiv beeinflussen (Baumert, Watermann & Schümer, 2003). Der Zugang zu diesen entwicklungsfördernden Ressourcen, wie etwa das Bereitstellen von Büchern oder Möglichkeiten der Förderung von Neugier und Interesse, welche nicht selten an die Verfügbarkeit finanzieller Mittel gebunden sind, ist für Familien mit niedrigerem SES erschwert, was Korrelationen zwischen SES und dem Zugang zu Bildungsgütern verdeutlichen (Bornstein & Bradley, 2003).

Auch die Beständigkeit der Umwelt scheint ein kritischer Aspekt zu sein, der zwischen Familien mit hohem und niedrigem SES differenziert. Denn während ein höherer SES mit einer homogenen, stabileren Umwelt einherzugehen scheint (Bradley, Corwyn, Burchinal, McAdoo & Garcia Coll, 2001; Brooks-Gunn & Duncan, 1997), zeigte die Studie von Ackerman, Brown und Izard (2004), dass unbeständige familiäre Gegebenheiten, wie etwa vermehrte Wohnort- oder Partnerwechsel, besonders häufig bei Familien mit niedriger sozialer Stellung auftreten. Hinsichtlich der Stabilität des SES ist anzumerken, dass allgemein steigende Scheidungsraten, häufigere Jobwechsel und damit verbundene Phasen der Arbeitslosigkeit, plötzliche Änderungen der finanziellen Situation und damit des familiären SES wahrscheinlicher machen (Jeynes, 2002a), auch wenn Gottfried, Gottfried, Bathurst, Guerin und Parramore (2003) die Stabilität des SES, über die gesamte Kindheit betrachtet, als sehr hoch einstufen. Der Aspekt der Stabilität ist deshalb von Relevanz, da von einem Einfluss von Veränderungen familiärer Strukturen und anderer Gegebenheiten auf die kindliche Entwicklung auszugehen ist (Jeynes, 2002a). So erweist sich also nicht nur der Zugang zu förderlichen Ressourcen für Kinder aus Familien mit niedrigem SES als erschwert (Bornstein & Bradley, 2003), auch die unstrukturierte Umwelt muss als Merkmal genannt werden, welches mit einem niedrigen SES assoziiert zu sein scheint und sich negativ auf die kindliche Entwicklung auswirken kann (Evans, Gonnella, Marcynyszyn & Salpekar, 2005).

Weiterhin ist der robuste Befund zu erwähnen, dass SES mit wichtigen Kriterien wie körperlicher und geistiger Gesundheit assoziiert ist (Johnson, Corley, Starr & Deary, 2011). Auch die empirisch abgesicherte Beziehung zu Intelligenz ist als relevantes Merkmal des SES zu nennen (Brody, 1992; Leventhal & Brooks-Gunn, 2000). Die berichteten Korrelationen liegen typischerweise zwischen $r = .30$ und $.40$ (Brody, 1992). Auch White (1982) berichtete eine Korrelation von $r = .33$ zwischen Intelligenzleistung des Kindes und elterlichem SES. Eine Studie an dreijährigen Zwillingen von Petrill und Deater-Deckard (2004) ergab einen Zusammenhang von $r = .47$ zwischen SES der Eltern und IQ der Kinder. Rowe, Vesterdal und Rodgers (1998) fanden Zusammenhänge von $r = .52 - .62$ zwischen Intelligenz und Bildung, sowie Werte zwischen $r = .26 - .38$ zwischen Intelligenz und Einkommen. Auch im Rahmen der PISA-Studie (OECD, 2001) wurde der Zusammenhang zwischen allgemeinen kognitiven Fähigkeiten und der elterlichen sozioökonomischen Stellung nachgewiesen ($r = .35$) (Brunner, 2006). Die Ursache dieser Beziehung sieht Jencks (1979) in der wechselseitigen Beeinflussung von IQ und SES. So kann ein niedriger SES, wie bereits dargestellt, den Zugang zu für die Entwicklung von Bildung förderlichen Ressourcen erschweren, etwa als Folge eines niedrigeren Einkommens. Dies wirkt sich wiederum kritisch auf die intellektuelle Entwicklung der Kinder aus (Jeynes, 2002b; Petrill et al., 2004). Diese Verknüpfung veranschaulicht eine Studie von Duncan, Brooks-Gunn und Klebanov (1994), in der ein Zusammenhang zwischen andauernder Armut innerhalb der ersten Lebensjahre und der Intelligenzentwicklung des Kindes nachgewiesen werden konnte. Da jedoch das mütterliche Verhalten sowie die Lernumwelt im Elternhaus die Beziehung zwischen IQ des Kindes und elterlichem Einkommen mediieren, rieten die Autoren davon ab, das Anheben des Einkommens als generelle Förderungs- oder Interventionsmöglichkeit zu forcieren (Duncan et al., 1994).

SES zeigt außerdem Zusammenhänge zu anderen wichtigen Faktoren, die die kognitive Entwicklung von Kindern beeinflussen können, wie beispielsweise Zugang zu Maßnahmen der Gesundheitsvorsorge oder zu gesunden Nahrungsmitteln (Bradley & Corwyn, 2002).

Ein höherer SES der Eltern und die damit verknüpfte förderlichere familiäre Umwelt kann sich nun positiv auf die Intelligenz des Kindes auswirken (Loehlin, 2000). Bei der kausalen Interpretation dieser Beziehung ist jedoch zu bedenken, dass sie zum Teil genetisch vermittelt wird (Bouchard & McGue, 2003). Denn elterliche Intelligenz und SES stellen stark assoziierte Faktoren dar, so dass ein höherer SES üblicherweise auch auf höhere elterliche Intelligenz schließen lässt, welche nun

wiederum an die Kinder vererbt wird (Loehlin, 2000). In diesem Zusammenhang ist zu erwähnen, dass verhaltensgenetische Studien zeigen konnten, dass im Laufe des Heranwachsens der Einfluss der familiären Umwelt auf den IQ stetig abnimmt, während die Bedeutung der genetischen Einflüsse ansteigt (Hart, Petrill, Deater-Deckard & Thompson, 2007; Plomin, DeFries, McClearn & Rutter, 1997, zitiert nach Loehlin, 2000). Insgesamt ist die Beziehung zwischen IQ und SES durch das Zusammenspiel von Gen-Umwelt-Korrelationen bzw. Gen-Umwelt-Interaktionen gekennzeichnet. So gibt es Hinweise darauf, dass bestimmte Gene, welche mit Intelligenz assoziiert sind, sich lediglich in Umwelten mit hohem SES entfalten (Gen-Umwelt-Interaktion). Außerdem geben Eltern immer gleichzeitig ihre Gene, welche mit Intelligenz verknüpft sind, sowie die durch den SES geprägte Umwelt weiter (Gen-Umwelt-Korrelation), so dass insgesamt eine wechselseitige Beeinflussung von SES und Intelligenz stattfindet (Deary et al., 2010).

Ein weiterer wichtiger Befund betrifft die Tatsache, dass sich Personen mit ähnlichem SES clustern, so dass sich auch die soziale Umwelt, in der ein Kind heranwächst, auf die kognitive Entwicklung auswirkt (Leventhal & Brooks-Gunn, 2003). Hierzu zählen etwa die Nachbarschaft oder die Wohnregion, in der ein Kind aufwächst, aber auch der Zugang zu institutionellen Ressourcen (beispielsweise Schulen), die je nach Wohnlage variieren kann. Auch Bornstein und Bradley (2003) weisen darauf hin, dass die Bedeutung der finanziellen Ressourcen mit zunehmenden Alter an Gewicht verliert, während die sozialen Netzwerke, die eine Familie umgeben, immer relevanter werden.

Neben der Bereitstellung entwicklungsfördernder Bedingungen, die an einen höheren SES gekoppelt sind, entfaltet der SES seine positive Wirkung auf die kindliche Entwicklung auch über den Zusammenhang mit bestimmten elterlichen Verhaltensweisen. So konnte eine Beziehung zwischen SES und ermutigendem Erziehungsverhalten nachgewiesen werden, sowie einer häufigeren Interaktion zwischen Mutter und Kind (Bornstein, Hahn, Suwalsky & Haynes, 2003). Von Bedeutung ist, dass besonders die mütterliche Bildung relevant zu sein scheint, was möglicherweise dadurch begründet werden kann, dass in den ersten Jahren häufiger die Mütter als primäre Bezugspersonen fungieren und mehr Zeit mit den Kindern verbringen als die Väter. Insgesamt scheint jedoch besonders das Bereitstellen einer stimulierenden Lernumgebung eine große Rolle für die Kompetenzentwicklung von Kindern zu spielen (Bradley & Corwyn, 2003).

Im Hinblick auf die Beziehung des SES zu den kognitiven Fähigkeiten gilt, dass die Höhe des Zusammenhangs durch die Operationalisierung der Konstrukte beeinflusst wird. Denn während kognitive Tests, die kristallisiertes Wissen aktivieren, stark mit Indikatoren des SES korrelieren, erweisen sich fluide Fähigkeitstests als unabhängiger. Daher werden sprachfreie Intelligenztests üblicherweise als kulturfair bezeichnet. Auch mit WM korreliert SES in der Tendenz nur gering. Während also beispielsweise standardisierte Vokabeltests deutlich mit SES korrelieren, weisen bestimmte WM-Aufgaben, die anstelle von Wörtern frei erfundene Nonwords einsetzen, keine Beziehung zum SES auf (Campbell, Dollaghan, Needleman & Janosky, 1997; EllisWeismer et al., 2000). Somit gilt besonders WM als weitestgehend unabhängig vom sozioökonomischen Hintergrund, da es eine grundlegende fluide kognitive Fähigkeit darstellt, die relativ wenig durch bereits angeeignetes Wissen konfundiert ist (Conway et al., 2002; Engle, Kane et al., 1999; Engle, Tuholski et al., 1999). Auch die Studie von Engel, Santos und Gathercole (2008) belegt die Unabhängigkeit verschiedener WM-Komponenten (PL und ZE) vom familiären SES. Die Tatsache, dass hingegen Vokabellernen hohe Korrelationen zum elterlichen SES aufweist, wird dadurch begründet, dass in Familien mit höherem SES Bildung tendenziell stärker wertgeschätzt wird und mehr mit den Kindern geredet und interagiert wird (z.B. Brody & Flor, 1998; Fuligni, 1997; Hoff, 2003; Huttenlocher, Haight, Bryk, Seltzer & Lyons, 1991).

Im folgenden Kapitel 1.5 wird das für diese Arbeit zentrale Konzept Schulerfolg beleuchtet, wobei ein Bezug zu allen bereits thematisierten Einflussfaktoren hergestellt wird.

1.5 Schulerfolg

Die Untersuchung der Bedingungsfaktoren schulischer Leistung wird als eine der ältesten und schwierigsten Fragestellungen der pädagogischen Psychologie verstanden (Helmke & Schrader, 2006). Schwierigkeiten ergeben sich bereits durch die nicht eindeutige Definition dessen, was unter Schulerfolg oder -leistung zu verstehen ist. So kann sich der Begriff auf die Einzelleistung eines Schülers beziehen oder auf die Leistungen auf Klassenebene. Die Frage der Operationalisierung von Schulerfolg ist ein weiterer Gesichtspunkt, durch den sich Unterschiede zwischen Studien ergeben, etwa ob Noten herangezogen werden oder standardisierte Leistungstests (Helmke & Schrader, 2006). Tendenziell wird im Zusammenhang mit Schulnoten häufiger von Schulerfolg gesprochen, während sich der Begriff der Schulleistung häufiger auf Studien im Kontext von Schulleistungstests bezieht. Studien unterscheiden sich außerdem darin, ob auf fachspezifischer Ebene erhoben wurde oder fächerübergreifend.

Da im deutschen Schulsystem ausschließlich Schulnoten als Leistungsindikatoren zum Einsatz kommen, welche auch über die Versetzung in die nächste Klassenstufe bzw. am Ende der vierten Klasse über die weiterführende Schulform entscheiden, wird Schulerfolg auch in der vorliegenden Arbeit mittels fachspezifischer Noten operationalisiert und nicht durch den Einsatz von Schulleistungstests. Zwar wird die Verwendung von Schulnoten als Leistungskriterium häufig aufgrund von Schwächen hinsichtlich der wichtigsten Gütekriterien Objektivität, Reliabilität und Validität bemängelt (Süß, 2001), da die Notengebung stets durch subjektive Einschätzungen des Lehrenden beeinflusst wird, welche durch die Vermengung von Beobachtung und Interpretation entstehen. Doch kritisch scheinen weniger regelhafte Verzerrungen zu sein, als vielmehr der Mangel eines einheitlichen Maßstabes über verschiedene Klassen hinweg, wodurch der Vergleich von Schulnoten erschwert wird (Süß, 2001). Die Tatsache, dass Schulnoten jedoch hoch signifikant mit den Leistungen in standardisierten Tests korrelieren (Hoge & Coladarci, 1989), und darüber hinaus deutliche Zusammenhänge zu Indikatoren von Studien- und Berufserfolg aufweisen (z.B. Robbins, Lauver, Le, Davis, Langley & Carlstrom, 2004) kann als Beleg für die Validität der Noten und damit als Rechtfertigung ihres Einsatzes als Leistungskriterium gewertet werden (Krumm, Ziegler & Bühner, 2008). Insgesamt kann „nach einer Phase heftiger Kritik“ (Süß, 2001, S. 121) an der

Notenvergabe in deutschen Schulen von „einer partiellen `Rehabilitation´“ (Süß, 2001, S. 121) gesprochen werden.

Neben den methodischen und konzeptionellen Gesichtspunkten die Definition des Kriteriums Schulerfolg betreffend, ist natürlich die Betrachtung der unterschiedlichen Wirkfaktoren auf das Konzept Schulerfolg relevant. Bislang fehlt ein allgemein anerkanntes übergreifendes Vorhersagemodell, was dem Umstand geschuldet ist, dass Schulerfolg dem Prinzip der multiplen Bedingtheit unterliegt, welches die Beachtung komplexer Wirkzusammenhänge erfordert und die Modellprüfung gleichzeitig erschwert (Helmke & Schrader, 2006). Doch es herrscht Einigkeit darüber, dass Schulerfolg nicht nur durch individuelle Schülerfaktoren beeinflusst wird, wie kognitive Fähigkeit, Vorwissen, Arbeitsverhalten, Persönlichkeit oder Motivation. Darüber hinaus erweisen sich besonders die Bereiche „Familie, Schule, Peers und Medien“ (Helmke & Schrader, 2006, S. 83) als relevant. Abbildung 2 veranschaulicht das komplexe Zusammenspiel der verschiedenen Einflussfaktoren.

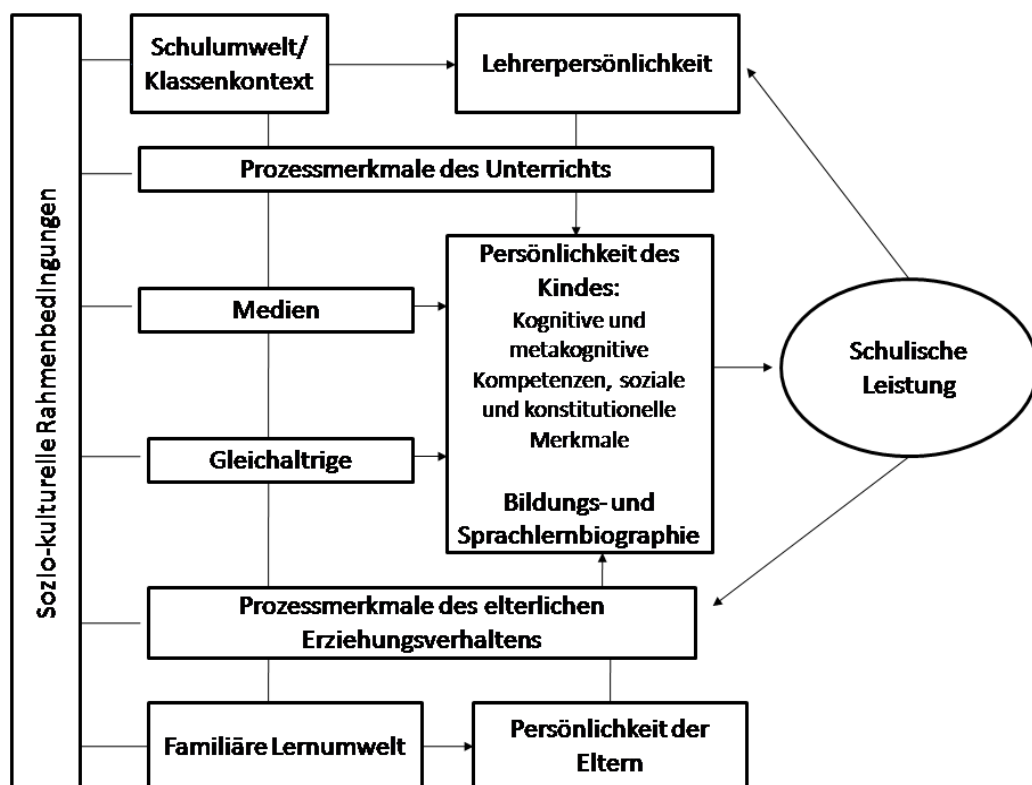


Abbildung 2: Modell der Bedingungsfaktoren schulischer Leistung (Helmke & Schrader, 2006)

Familiäre Einflüsse beziehen sich dabei nicht nur auf die Persönlichkeit der Eltern oder deren Erziehungspraktiken. Auch die in einer Familie herrschenden Lernbedingungen oder allgemeiner die sozio-kulturellen Rahmenbedingungen wirken dem Modell folgend indirekt auf die schulische Leistung eines Kindes.

Selbstverständlich übt auch der Lehrer selbst durch seine Persönlichkeit und den damit verbundenen individuellen Unterrichtsstil einen Einfluss aus. Insgesamt ist von Interaktionseffekten zwischen Schüler-, Schul- und anderen Umweltvariablen auszugehen (Helmke & Weinert, 1997; Süß, 2001). Süß (2001) folgerte aus den Befunden der beiden bedeutenden Meta-Analysen von Fraser, Walberg, Welch und Hattie (1987) bzw. Wang, Haertel und Walberg (1993), dass zwar die kognitiven Merkmale des Schülers wie Vorwissen oder Intelligenz als wichtigste Einflussfaktoren auf die Schulleistung zu betrachten sind. Allerdings übt der Umgang des Lehrers mit der Klasse einen ebenfalls nicht zu unterschätzenden Einfluss aus. Schließlich bilden nicht-kognitive Merkmale des Kindes, sowie Verhaltensweisen der Eltern und andere familiäre Gegebenheiten eine weitere Gruppe relevanter Wirkfaktoren (Süß, 2001).

Aufgrund des Wechselspiels dieser verschiedenen Einflussgrößen ist nach Helmke und Schrader (2006) zu beachten, dass die isolierte Betrachtung einer einzelnen Komponente immer zu verzerrten Schlussfolgerungen führen kann und daher eine möglichst umfassende Berücksichtigung verschiedener Faktoren stets erstrebenswert ist. So beschränkt sich die vorliegende Arbeit nicht auf die vergleichende Betrachtung der kognitiven Variablen Intelligenz und WM, sondern schließt weiterhin non-kognitive Individualfaktoren in die Analysen ein (FSK, IM), sowie den sozioökonomischen Hintergrund als Familienvariable. Die angesprochenen Determinanten des Schulerfolgs werden nun tiefergehend betrachtet.

1.5.1 Kognitive Prädiktoren und Schulerfolg

1.5.1.1 Die Bedeutung der Intelligenz als Prädiktor des Schulerfolgs

Intelligenz oder allgemeine kognitive Fähigkeit gilt als bedeutendster und validester Prädiktor schulischer Leistungen (Kuncel, Hezlett & Ones, 2004). Vor dem Hintergrund, dass der Begriff Intelligenz vor allem mit Lernfähigkeit in Verbindung gebracht wird, verwundert dies nicht (Spinath, Spinath, Harlaar & Plomin, 2006). Der

Befund, dass hochaggregierte Maße der allgemeinen Intelligenz und Schulleistung im Mittel mit $r = .50$ korreliert sind, kann als empirisch abgesichert bezeichnet werden (Gustafsson & Undheim, 1996). So ergab die Studie von Fraser et al. (1987), welche verschiedene meta-analytische Befunde zusammenfasste, dass die Korrelationen zwischen Werten von $r = .43$ (Steinkamp & Maehr, 1983) bis $r = .51$ (Hattie & Hansford, 1982) streuten. Leider differenzieren diese Studien nicht zwischen Schulfach, Schultyp oder Subkonstrukten der Intelligenz (Süß, 2001). Spezifische Intelligenzfaktoren können jedoch über g hinaus einen inkrementellen Beitrag bei der Vorhersage spezifischer Fähigkeiten leisten (Gustafsson & Balke, 1993). In diesem Zusammenhang ist eine aktuelle Studie von Vock, Preckel und Holling (2011) erwähnenswert. Die Autoren untersuchten eine Stichprobe von annähernd 1500 Schülern der siebten bis zehnten Klassenstufe und demonstrierten, dass Gedächtniskomponenten wichtige Prädiktoren von Schulnoten darstellen, und zwar über die Reasoningfähigkeit hinaus. Daher fordern die Autoren weitere Studien, die im Zuge multimodaler Erhebungen, sowohl einfache als auch komplexe kognitive Fähigkeiten im Hinblick auf die Vorhersagekraft von Schulerfolg untersuchen (Vock et al., 2011).

Allgemein gilt, dass die Höhe der Korrelation zwischen Intelligenz und Schulerfolg zum einen durch die Art der Operationalisierung von Intelligenz- und Schulleistung bzw. -erfolg beeinflusst wird (einzelne Schulnoten versus Schulleistungstest etc.), aber auch durch Alter, Schultyp und den sozialen Hintergrund der Probanden (Gustafsson & Undheim, 1996; Heller, 1997). So finden sich im Grundschulalter höhere Korrelationen als in weiterführenden Schulen, was als Folge der Varianzeinschränkung aufgrund der im Sekundarschulalter einsetzenden Selektionsprozesse zu sehen ist (Gustafsson & Undheim, 1996; Humphreys & Stark, 2002). Die starke Verbindung zwischen Schulerfolg und allgemeiner kognitiver Fähigkeit erklärt sich durch die Fähigkeit intelligenterer Kinder, Regeln und Zusammenhänge besser und schneller zu erkennen und Probleme effizienter zu lösen. Außerdem ist anzunehmen, dass sie über effizientere Wissensstrukturen verfügen, welche sich wiederum positiv auf den weiteren Wissenserwerb auswirken (Helmke & Schrader, 2006). Doch auch wenn kognitive Fähigkeiten mit einer Varianzaufklärung von 25% bis 45% (Grube & Hasselhorn, 2006) als mächtigste Prädiktoren gelten (Heller, 1997) bleibt Schulerfolg dennoch multipel determiniert (Helmke & Schrader, 2006), weshalb stets auch andere Einflussgrößen relevant sind (Neisser et al., 1996). Außerdem weisen Helmke und Schrader (2006) darauf hin, dass die jahrelange Fixierung auf kognitive Prädiktoren zu einem verzerrten Blick auf

andere wichtige Einflussgrößen geführt hat, weshalb beispielsweise die Relevanz des Vorwissens lange Zeit unberücksichtigt blieb, obwohl es Hinweise darauf gibt, dass dieses sogar einen größeren Beitrag zur Vorhersage leisten kann, als die allgemeine Intelligenz (Weinert & Helmke, 1995).

Die Sichtung der Zusammenhänge zwischen Intelligenz und den verschiedenen Formen schulischer Leistung ergibt, dass die Korrelation zu standardisierten Leistungstests höher ausfällt ($r = .61$ bis $.90$) als zu Schulnoten (Deary et al., 2007; Rindermann, 2006). Dies kann als Hinweis auf geteilte Methodenvarianz verstanden werden (Steinmayr & Spinath, 2009). Andererseits stellen Schulnoten nie reine Leistungsindikatoren dar, sondern beinhalten immer auch verhaltensbezogene und motivationale Aspekte, wie etwa das soziale Verhalten während des Unterrichts, welche bei der Notenvergabe Berücksichtigung finden (Steinmayr & Spinath, 2009). Allgemein führen fachspezifische Erhebungen üblicherweise zu höheren Korrelationen (Amthauer, Brocke, Liepmann & Beauducel, 2001), auch wenn sich nicht in allen Studien ein Vorteil einer domänen-spezifischen Intelligenzmessung gegenüber einer allgemeinen ergab (Spinath et al., 2006). Der stärkste Zusammenhang zwischen Schulnote und allgemeiner Intelligenz findet sich im Fach Mathematik, wo Korrelationen um $r = .40$ auftreten (Amthauer et al., 2001). Auch Krumm et al. (2008) berichten höhere Werte in der naturwissenschaftlichen Domäne ($r = .35$), im Vergleich zur sprachlichen Domäne ($r = .18$).

1.5.1.2 Die Bedeutung von Arbeitsgedächtnis als Prädiktor des Schulerfolgs

Neben der allgemeinen Intelligenz spielt besonders die Arbeitsgedächtniskapazität eine wichtige Rolle als kognitiver Prädiktor schulischer Leistung. Nach Miyake und Shah (1999) umfasst die Instanz WM „mechanisms or processes that are involved in the control, regulation, and active maintenance of task-relevant information in the service of complex cognition“ (S. 450) und ist daher besonders für die Ausführung höherer kognitiver Leistungen bedeutsam. WM gilt als nur wenig durch Einflüsse der sozialen Umwelt wie den sozioökonomischen Status der Eltern oder Vorerfahrung beeinflussbar, weshalb es als relativ unverfälschter Indikator des kindlichen Lernpotenzials angesehen wird (Alloway, 2009; Alloway et al., 2004). Zwischenzeitlich mehren sich Studien, welche die Beziehung zwischen WM und

Lernfähigkeit ebenso empirisch untermauern, wie die prädiktive Kraft dieses Konstrukts hinsichtlich schulischer Leistung (Alloway et al., 2005; Gathercole, Brown, & Pickering, 2003; Gathercole, Pickering, Knight & Stegmann, 2004). Bereits 1980 berichteten Daneman und Carpenter eine Korrelation von $r = .59$ zwischen der Leistung im Verbalteil des SAT (Scholastic Aptitude Test) und Arbeitsgedächtnisaufgaben. In den letzten Jahren lieferten besonders die Studien von Gathercole und Kollegen wichtige Hinweise für die Bedeutung des WMs als Prädiktor schulischer Leistung. So zeigte sich bereits im Alter von sieben Jahren ein Bezug zwischen schwacher Leistung in den Fächern Englisch und Mathematik und Einschränkungen der Arbeitsgedächtniskapazität (Gathercole & Pickering, 2000b; Gathercole et al., 2003). Außerdem konnte ein Zusammenhang zwischen schwacher Arbeitsgedächtnisleistung in verbalen Spannaufgaben und Aufmerksamkeitsproblemen im Sinne einer verringerten Aufmerksamkeitsspanne oder leichter Ablenkbarkeit festgestellt werden (Gathercole et al., 2008). Die diagnostische Bedeutung des WMs im Hinblick auf die Erkennung von Lernschwierigkeiten bestätigte auch die aktuelle Meta-Analyse von Carretti, Borella, Cornoldi und De Beni (2009). Diese zeigte, dass durch Messung der Arbeitsgedächtnisleistung Lernschwierigkeiten im Bereich des Leseverständnisses aufgedeckt werden können. Hierzu eigneten sich besonders Aufgaben, die Kontroll- bzw. Aufmerksamkeitsprozesse erfordern (Carretti et al., 2009).

Insgesamt liegen für verschiedene Kernbereiche schulischer Leistung bzw. für Fähigkeiten, die als Voraussetzung guter schulischer Leistung verstanden werden, Evidenzen für die Bedeutung des WMs vor, wie beispielsweise Lesefähigkeit (de Jong, 1998; Swanson, 1994), Rechenkompetenz (Bull & Scerif, 2001; De Smedt et al., 2009; McLean & Hitch, 1999; Passolunghi & Siegel, 2001), Hörverstehen (Daneman & Carpenter, 1980), Leseverständnis (Turner & Engle, 1989) oder Wortschatz (Daneman & Green, 1986).

Aktuelle Studien gehen der Frage nach, auf welchen spezifischen Komponenten und Prozessen diese Zusammenhänge basieren. So wurde in der Studie von Clark, Pritchard und Woodward (2010) die prädiktive Validität spezifischer exekutiver Funktionen (komplexes Problemlösen, Shifting und Inhibition) bezüglich der Mathematikleistung von Grundschulern untersucht. Dabei fand sich, dass die im Alter von vier Jahren gemessenen exekutiven Funktionen gemeinsam die Leistung im Woodstock-Johnson Test of Achievement ein Jahr nach Schuleintritt vorhersagen konnten, auch unter Kontrolle des SES, der allgemeinen kognitiven Fähigkeiten und

der Leseleistung. Allerdings lag der inkrementelle Beitrag bei lediglich 3%, so dass eine Replikation dieses Befundes erstrebenswert wäre. Außerdem existieren Hinweise, dass bei Vorschulkindern und Kindern unter acht Jahren inhibitorische Fähigkeiten mit der Mathematikleistung in Verbindung stehen (Bull & Scerif, 2001; Espy et al., 2004). Konsistente Befunde fanden St. Clair-Thompson und Gathercole (2006) für elfjährige Kinder. Auch Shifting konnte bei Kindern über sieben Jahren die Leistung vorhersagen (Bull, Johnston & Roy, 1999; van der Sluis, van der Leij & de Jong, 2004). Exekutive Funktionen erwiesen sich jedoch auch in allen wichtigen sprachlichen Lernbereichen als äußerst relevant (Bull et al., 2008). Während inhibitorische Fähigkeiten besonders beim Vokabellernen oder Lesen eine Rolle zu spielen scheinen, ist Shifting beim Kopfrechnen und Schreiben von großer Wichtigkeit (St. Clair-Thompson & Gathercole, 2006).

Ob nun die Zusammenhänge zwischen WM und schulischer Leistung durch domänen-spezifische Prozesse verursacht werden, ist umstritten. Zwar fanden Jarvis und Gathercole (2003) bei Elf- bis Vierzehnjährigen domänen-spezifische Assoziationen, da die stärksten Zusammenhänge zwischen verbaler WM-Leistung und der Englischleistung bzw. zwischen visuell-räumlichem WM und der Leistung in Mathematik gefunden wurden. Doch in einer Folgestudie konnten diese fächerspezifischen Ergebnisse nicht bestätigt werden (St. Clair-Thompson & Gathercole, 2006), was für eine generelle Relevanz exekutiver Funktionen für alle Lernprozesse spricht. Einen weiteren Hinweis auf domänen-spezifische Zusammenhänge lieferte die Studie von Vock und Holling (2008). Denn hier erwies sich eine verbal-numerische Arbeitsgedächtnisfacette als bester Prädiktor der Deutschnote, die sogar die prädiktive Validität der fluiden Intelligenz überschritt, während die Mathematiknote besonders gut durch die visuell-räumliche Arbeitsgedächtniskomponente vorhergesagt werden konnte.

De Smedt et al. (2009) fassen zusammen, dass zwar die zentrale Rolle der ZE im Baddeley'schen Verständnis unbestritten ist, jedoch die Bedeutung der untergeordneten Subsysteme nach wie vor unklar bleibt. In ihrer Studie erwies sich die ZE sowohl in der ersten als auch in der zweiten Klasse als signifikanter Prädiktor der Mathematikleistung. Der VSSP konnte jedoch lediglich in Klasse 1 einen eigenständigen Anteil aufklären, nicht in Klasse 2. Umgekehrt verhielt es sich mit der PL. Die Autoren sehen diese Befunde durch entwicklungsbedingte Veränderungen innerhalb des Arbeitsgedächtnissystems verursacht, da Kinder mit zunehmenden Alter verstärkt auf verbale Lösungsstrategien zurückgreifen und somit die Rolle des VSSP geschwächt wird (De Smedt et al., 2009). Bezüglich der Rolle der PL zeigten

zwar die Studien von Hecht, Torgesen, Wagner und Rashotte (2001) sowie von Swanson und Kim (2007) einen eindeutigen Zusammenhang zur Leistung in Mathematik bei Grundschulern, allerdings ergab sich dieser Zusammenhang bei Gathercole und Pickering (2000b) lediglich, wenn die Leistung der ZE nicht kontrolliert wurde. In der Untersuchung von Geary, Hamson und Hoard (2000) zeigte sich außerdem, dass Kinder mit Lernproblemen in Mathematik keine Beeinträchtigungen der PL aufwiesen, so dass die Bedeutung der PL für mathematische Fähigkeiten nicht abschließend geklärt ist.

Auch die Befunde zur Rolle des VSSP sind widersprüchlich. Zwar existieren auch hier Hinweise auf einen Bezug zur Leistung in Mathematik im Grundschulalter (Holmes & Adams, 2006), doch wie bereits erwähnt zeichnet sich ab, dass der Einfluss des VSSP mit steigendem Alter abnimmt, da ältere Kinder zunehmend verbale Strategien anwenden (De Smedt et al., 2009; Holmes & Adams, 2006).

Schlussendlich bleibt festzustellen, dass WM alle Bereiche des Lernens beeinflusst und deshalb von großem diagnostischem Wert ist, wenn es darum geht, Lernschwierigkeiten zu vermeiden bzw. frühzeitig aufzudecken. So ist für Kinder mit Arbeitsgedächtnisdefiziten nicht nur das Merken und Ausführen von Instruktionen während des Unterrichts problematisch. Auch Kernkompetenzen wie Kopfrechen oder das Aufschreiben von im Kopf formulierten Texten stellen enorme Herausforderungen dar (St. Clair-Thompson & Gathercole, 2006). Die Zusammenhänge zwischen WM und Schulleistung konnten für verschiedene Nationalitäten, Schulsysteme und Klassenstufen belegt werden, so dass die Beziehung als empirisch abgesichert bezeichnet werden kann.

1.5.1.3 Vergleich der prädiktiven Validität von Arbeitsgedächtnis und Intelligenz

Wenige Studien haben sich bisher mit dem Vergleich der prädiktiven Kraft von WM und Intelligenz im Hinblick auf Schulnoten auseinandergesetzt. Viele Studien fokussierten entweder auf Lernschwierigkeiten (Alloway, 2009), untersuchten ältere Stichproben (Krumm, Ziegler & Bühner, 2008) oder verwendeten Leistungstests oder ähnliche Kriteriumsoperationalisierungen anstelle von Noten (z.B. Gathercole et al., 2004). Außerdem erwiesen sich die bisherigen Ergebnisse als nicht konsistent. So ergab beispielsweise die Studie von Luo et al. (2006), dass in einer Gruppe von

Sech- bis Neunzehnjährigen die Leistung in einer Arbeitsgedächtnisbatterie einen inkrementellen Beitrag (ca. 15%) über die Leistung im WISC-R hinaus vorhersagen konnte. Allerdings wurden hier keine Schulnoten betrachtet. Die bereits erwähnte Untersuchung von Vock und Holling (2008), an der unter anderem Grundschüler der vierten bis sechsten Klassenstufe teilnahmen, konnte die Befunde von Luo et al. (2006) untermauern, da die Leistung in Arbeitsgedächtnisaufgaben sowohl in Mathematik, als auch in Deutsch und Sachkunde inkrementell zur Varianzaufklärung in den jeweiligen Noten beitrug. Die inkrementelle Validität variierte dabei je nach Domäne und Subtest. Besonders deutlich fielen die Ergebnisse im Fach Deutsch aus, denn hier konnte WM 22% über gf hinaus vorhersagen, 25% über numerische Intelligenz und 8% über verbale Intelligenz. In Mathematik klärte WM lediglich 7% über gf hinaus auf, allerdings 15% über verbale sowie 10% über numerische Intelligenz. Die Studie von Rohde und Thompson (2007), in der eine studentische Stichprobe untersucht wurde, ergab jedoch den konträren Befund, dass der inkrementelle Beitrag des WMs über die allgemeine kognitive Fähigkeit hinaus sich als nicht signifikant erwies. Allerdings wurde WM hier lediglich anhand der *Operation Span Task* operationalisiert. Auch in der Studie von Krumm et al. (2008) vermochte WM es nicht, über Reasoning hinaus weitere Varianzanteile aufzuklären. Allerdings belegte die Tatsache, dass die um WM bereinigte Reasoning-Varianz ebenfalls keinen signifikanten Beitrag mehr leisten konnte, die Wichtigkeit der gemeinsamen Varianzanteile von WM und Intelligenz (Krumm et al., 2008). Da sich diese Studie jedoch ebenfalls auf eine Studentenstichprobe bezog, ist die Übertragbarkeit der Befunde auf eine jüngere Stichprobe ungeklärt.

Alloway (2009) verglich die prädiktive Kraft von WM und Intelligenz zur Vorhersage des Lernerfolges nach zwei Jahren in einer Stichprobe von 64 Kindern mit Lernschwierigkeiten im Alter von sieben bis elf Jahren. Hier zeigte sich das umgekehrte Muster, dass Intelligenz keinen Beitrag über WM und Vorwissen hinaus leisten konnte. Allerdings erscheint die Generalisierbarkeit dieser regressionsanalytischen Befunde aufgrund der sehr kleinen Stichprobe (N = 37 zu T2) als fraglich.

Dass WM für sich genommen hinsichtlich seiner prädiktiven Validität durchaus auch bei Kindern mit der allgemeinen Intelligenz mithalten kann, zeigte die Studie von de Jong und Das-Smaal (1995), in der Viertklässler untersucht wurden. Hier wurde ein Zusammenhang von $r = .73$ zwischen einem latent geschätzten Arbeitsgedächtnis- und einem Schulleistungsfaktor gefunden, wobei dieser Zusammenhang sich nicht

signifikant von dem für die Beziehung zwischen Intelligenz und Schulleistung geschätzten Pfad ($r = .86$) unterschied.

Weitere Erkenntnisse bezüglich der Rolle von WM und Intelligenz für die Leistung in Mathematik lieferte die Pfadanalyse von Passolunghi, Mammarella und Altoè (2008). Hier fand sich ein signifikanter Zusammenhang zwischen WM und der Leistung in einem Mathematikleistungstest, sowohl in der ersten als auch in der zweiten Klasse, während die Leistung im *performance IQ*-Subtest des WISC-R lediglich in der ersten Klasse einen direkten Effekt zeigte. In der zweiten Klasse konnte dieser Faktor die Mathematikleistung nicht mehr vorhersagen, was mit Unterschieden im Testmaterial des Leistungstests begründet wurde (Passolunghi et al., 2008). In einer früheren Studie (Passolunghi, Vercelloni & Schadee, 2007) wurde die Beziehung zwischen *performance IQ* und Mathematikleistung durch WM und Zählfähigkeit mediert, allerdings handelte sich dabei um latente Modellierungen, während die Studie von 2008 Pfadanalysen berichtete. In der Studie von 2008 wurde jedoch der Einfluss des verbalen IQ durch WM mediert, so dass sich insgesamt eine hohe Relevanz des WMs abzeichnete. Die Autoren kamen zusammenfassend zu dem Schluss, dass sich die bedeutende Rolle des WMs in Einklang zu neuropsychologischen Studien befindet, die belegen, dass Problemlösen ähnliche Hirnareale im präfrontalen Kortex aktiviert, wie das Ausführen von Rechenoperationen (Gruber, Indefrey, Steinmetz & Kleinschmidt, 2001; Postle et al., 2006).

Auch wenn sich also tendenziell abzeichnet, dass die prädiktive Validität von WM und Intelligenz bezüglich der Vorhersage von Schulleistung ähnlich hoch ausfällt und dass besonders der hohe Anteil an geteilter Varianz von Relevanz ist, ist nach wie vor offen, ob diese Zusammenhänge für alle Altersgruppen, Domänen und Arten von Leistungskriterien gelten. Auch ist unklar, durch welche Prozesse und Strukturen spezifische Anteile an Kriteriumsvarianz aufgeklärt werden können, so dass es eine Weiterverfolgung dieser Fragestellungen sicherlich wertvoll ist.

1.5.2 Non-kognitive Prädiktoren und Schulerfolg

Die Frage inwiefern non-kognitive Variablen zur Vorhersage schulischer Leistungen beitragen können, hat eine lange Forschungstradition in der pädagogischen Psychologie. Neben der Bedeutung von Persönlichkeitsvariablen, wie beispielsweise Gewissenhaftigkeit (Laidra, Pullmann & Allik, 2007), werden besonders motivationale

Variablen hinsichtlich ihrer prädiktiven Validität untersucht. Motivation ist nicht zuletzt deshalb zentraler Forschungsgegenstand, weil davon ausgegangen wird, dass es sich hierbei um einen Faktor handelt, der durch äußere Gegebenheiten zu beeinflussen ist, und zwar besser als beispielsweise kognitive Fähigkeiten. Dies weckt die Hoffnung, aus den sich ergebenden Befunden Interventionsmöglichkeiten zur Steigerung der motivationalen Zustände von Schülern ableiten und damit positiv auf die schulischen Leistungen einwirken zu können (Spinath et al., 2006).

Während die Erfassung der Persönlichkeit im frühen Schulalter jedoch an Grenzen stößt, können bereits Grundschüler Selbsteinschätzungen hinsichtlich motivationaler Tendenzen vornehmen. So wurde auch in der vorliegenden Studie das Augenmerk besonders auf motivationale Aspekte gelegt. Dass Motivation sich auf die schulische Leistung von Kindern und Jugendlichen auswirkt, gilt als unbestritten (Murphy & Alexander, 2006). Als unklar muss jedoch das Ausmaß des Einflusses bewertet werden, denn auch wenn sich zahlreiche Belege für den signifikanten Zusammenhang von Motivation und Schulerfolg finden, scheint die Stärke der Beziehung geringer zu sein als erhofft bzw. erwartet (Sauer, 2006). Die Metaanalyse von Fraser et al. (1987) ergab beispielsweise lediglich einen Zusammenhang von $r = .29$ zwischen Motivation und Leistungskriterien. Außerdem fanden sich nur schwach positive Zusammenhänge zwischen Annäherungs-Leistungszielen und Lernzielen zu Schulerfolg (z.B. Greene & Miller, 1996), obwohl Lernziele mit tieferer Informationsverarbeitung und profunderem Umgang mit Lerninhalten in Verbindung gebracht werden (Ames & Archer, 1988; Meece, Blumenfeld & Hoyle, 1988). Zwischen Vermeidungslernzielen, Arbeitsvermeidung und Schulerfolg zeichnet sich tendenziell ein negativer Zusammenhang ab (Elliot, McGregor & Gable, 1999; Spinath et al., 2002). Die bereits erwähnte Problematik der zahlreichen Operationalisierungsmöglichkeiten motivationaler Konstrukte erschwert insgesamt die direkte Vergleichbarkeit der Ergebnisse. Deshalb werden im Folgenden besonders die Befunde fokussiert, die sich auf die für diese Arbeit zentralen Aspekte subjektiver Fähigkeitsüberzeugungen und Aufgabenwerte beziehen.

1.5.2.1 Die Bedeutung des Fähigkeitsselbstkonzepts als Prädiktor des Schulerfolgs

Mit moderaten Korrelationen von $r = .40$ bis $.60$ kann der Einfluss von Fähigkeitsüberzeugungen auf den schulischen Erfolg als substantiell bewertet werden (Guay et al., 2003; Marsh, Parker & Smith, 1983; Skaalvik & Valas 1999). Dabei fallen die Zusammenhänge üblicherweise höher aus, wenn eine fachspezifische Betrachtung erfolgt und wenn Noten statt Leistungstests als Kriterium herangezogen werden (Schicke & Fagan, 1994; Valentine, DuBois & Cooper, 2004). Außerdem steigern sich die Zusammenhänge mit zunehmenden Alter und Fähigkeitsniveau (Helmke & Weinert, 1997). Das subjektiv erlebte Fähigkeitsbild ist deshalb von Relevanz für die schulische Leistung, weil es beeinflusst, in welchem Maße leistungsrelevantes Verhalten initiiert und aufrechterhalten wird (Helmke & Weinert, 1997). So führen gering ausgeprägte Fähigkeitsüberzeugungen üblicherweise zur Entwicklung von Vermeidungsstrategien, durch die der Schüler versucht sich vor Misserfolgen zu schützen, um den Selbstwert zu stabilisieren (Helmke & Weinert, 1997).

Insgesamt gelten Fähigkeitsüberzeugungen als stärkste motivationale Prädiktoren schulischer Leistung (Helmke & Weinert, 1997; Wigfield, Tonks & Klauda, 2009). So fand Eccles (1984) positive Korrelationen zwischen fachspezifischem FSK und Schulnoten in Englisch und Mathematik, wobei sich die Beziehung in Englisch als stärker erwies im Vergleich zu Mathematik. In der Studie von Schicke und Fagan (1994), in welcher der GPA (grade point average) mittels einer globalen Fähigkeitseinschätzung vorhergesagt wurde, ergaben sich ebenfalls moderate Zusammenhänge. Eine aktuellere Studie von Schilling, Sparfeldt und Rost (2006) lieferte im Fach Deutsch einen Zusammenhang von $r = .47$ zwischen fachspezifischem FSK und Note, sowie einen Wert von $r = .57$ in Mathematik.

Als Kritik an Studien zum Zusammenhang zwischen FSK und Schulnoten wird häufig angeführt, dass unklar bleibt, inwiefern Fähigkeitsüberzeugungen lediglich internalisierte Repräsentation der vorherigen Noten darstellen. So wäre es denkbar, dass sich nur dann gesteigerte Fähigkeitsüberzeugungen entwickeln, wenn im Vorfeld auch bessere schulische Leistungen erzielt wurden. Zahlreiche Längsschnittstudien befassten sich somit mit der Frage der kausalen Ordnung von Leistung und Fähigkeitsüberzeugen. Während das *Self-Enhancement-Modell* postuliert, dass die schulische Leistung durch die jeweilige Fähigkeitsüberzeugung

determiniert wird, vertreten Anhänger des *Skill-Development-Ansatzes* die Auffassung, dass sich Fähigkeitsüberzeugungen als Konsequenz vorheriger Leistungen herausbilden (Calsyn & Kenny, 1977). Zwischenzeitlich herrscht jedoch Konsens, dass diese beiden Ansätze aufgrund der dichotomen Betrachtung von FSK als entweder Ursache oder Konsequenz die Zusammenhänge zu stark vereinfachen (Moschner & Dickhäuser, 2006). Folglich wird aktuell, besonders aufgrund der Arbeiten von Marsh und Kollegen, von reziproken Effekten ausgegangen. Denn Marsh und Kollegen (2005) konnten nachweisen, dass sowohl die vorherige Leistung nachfolgende Fähigkeitsüberzeugungen beeinflusst, als auch umgekehrt vorherige Fähigkeitseinschätzungen auf die nachfolgende Leistung einwirken. Besonders hervorzuheben ist dabei, dass der Einfluss des FSK auf nachfolgende Leistungen deutlicher ausfiel als der umgekehrte Zusammenhang. Auch die Metaanalyse von Valentine et al. (2004), die sich ausschließlich auf längsschnittliche Studien bezog, ergab, dass sich unter Kontrolle der initialen Leistung positive Selbsteinschätzungen günstig auf spätere Leistungen auswirken, auch wenn mit einem standardisierten Pfadgewicht von $\beta = .08$ von einem geringen Effekt gesprochen werden muss. Die Studie von Guay und Kollegen (2008) bestätigte ebenfalls das Modell reziproker Effekte für die Klassenstufen 2, 3 und 4. Doch es existieren auch gegensätzliche Befunde. So fanden Helmke und van Aken (1995) in den Klassenstufen 2 bis 4 keinen nennenswerten Einfluss des vorherigen Selbstkonzepts auf nachfolgende Mathematikleistung, so dass sie schlussfolgerten „during elementary school self-concept is mainly a consequence of cumulative achievement-related success and failure and that it does not have a significant impact on later achievement, neither on marks nor on test performance“ (S. 635). Insgesamt bedarf es also weiterer längsschnittlicher Studien, um die kausale Ordnung empirisch abzusichern.

1.5.2.2 Wertkomponenten als Prädiktoren des Schulerfolgs

Hinsichtlich der Bedeutung subjektiver Wertkomponenten wie Interesse oder intrinsischer Motivation im Hinblick auf die Vorhersage von schulischer Leistung, zeichnet sich trotz einer gewissen Variabilität der Befunde ab, dass die prädiktive Kraft dieser Faktoren im Vergleich zu den Fähigkeitsüberzeugungen abfällt (Spinath, Freudenthaler & Neubauer, 2010; Steinmayr & Spinath, 2007). Dennoch wird von einer positiven Rückkopplung ausgegangen, denn IM und Interesse führen zu einer ausdauernden Beschäftigung mit einem bestimmten Thema, welche sich

anschließend über Lernerfolge in gesteigertem Interesse widerspiegelt. Da jedoch laut Eccles et al. (1984) die Ausprägung an intrinsischer Motivation besonders dann eine Wirkung entfalten kann, wenn es darum geht, Entscheidungen zu treffen, beispielsweise hinsichtlich der Kurswahl, erscheinen die Befunde, dass sich die Zusammenhänge zu Leistungskriterien als niedriger erweisen im Vergleich zu Fähigkeitseinschätzungen, als theoriekonform. Doch trotz geringerer Assoziationen zu Schulleistung kann nicht von Irrelevanz gesprochen werden. So schlussfolgert Gottfried (1990) aus den Befunden zweier längsschnittlicher Studien, dass IM aufgrund der nicht mit Intelligenz, SES oder Vorleistung geteilten Varianzanteile durchaus als valider Prädiktor zu bezeichnen ist. Auch die Metaanalyse von Schiefele und Schreyer (1994) ergab niedrige, aber dennoch positive Zusammenhänge zwischen IM und Schulnoten bzw. Leistungstestergebnissen. Etwa 4% der Kriteriumsvarianz konnten durch IM aufgeklärt werden. Insgesamt können die Zusammenhänge zwischen Interesse, IM und Schulerfolg auf fachspezifischer Ebene mit Werten von $r < .30$ als schwach bis moderat klassifiziert werden (Gottfried, 1990; Lloyd & Barenblatt, 1984). Dabei legt die längsschnittliche Studie von Köller, Baumert und Schnabel (2001) eine wechselseitige Beeinflussung von Interesse und Schulleistung nahe, vergleichbar mit der Beziehung zwischen FSK und Schulerfolg. Allerdings bezog sich die Studie von Köller et al. (2001) ausschließlich auf Gymnasiasten, was die Generalisierbarkeit der Befunde einschränkt. Generell gilt, dass die Inkonsistenz der Befunde teilweise auf die Betrachtung von Stichproben unterschiedlichen Alters und die inkonsistente Kriteriumsoperationalisierung zurückzuführen ist.

Die Bedeutung von IM im schulischen Kontext beschränkt sich jedoch nicht auf den direkten Einfluss auf die Schulleistung. Denn IM korrespondiert stärker als extrinsische Lernmotivation mit Lernstrategien, welche eine tiefere Verarbeitung ermöglichen (Schiefele & Schreyer, 1994; Walker, Green & Mansell, 2006). Interessant ist der Befund, dass scheinbar besonders Kinder mit niedrigerer Intelligenz von ausgeprägter IM profitieren. Denn eine Studie von Tzuriel und Klein (1983) ergab, dass in einer Schülerstichprobe, in der Gruppen von hoch, mittel und niedrig intelligenten Kindern unterschieden wurden, zwar stets die intrinsisch motivierteren Schüler bessere Schulleistungen erzielten. Der Unterschied zu extrinsisch motivierten Schülern war jedoch in der Subgruppe der niedrig Intelligenten besonders deutlich. Die Autoren vermuteten, dass dieser Befund damit zusammenhängen könnte, dass intrinsisch Motivierte sich gerne herausfordernden Aufgaben stellen, was bei wenig intelligenten Kindern von größerem Nutzen sein

könnte, weil diese wahrscheinlich üblicherweise Herausforderungen meiden (Schiefele, 2009).

1.5.2.3 Inkrementelle Validität motivationaler Faktoren

Die Tatsache, dass subjektive Fähigkeitsüberzeugungen und Werteinschätzungen für sich genommen mit Schulleistung assoziiert sind, reicht jedoch nicht aus, um die komplexen Wirkmechanismen vollständig abzubilden, die die jeweilige Leistung eines Schülers bedingen. Denn da es sowohl zwischen FSK und kognitiven Fähigkeiten Überlappungen gibt (Ackerman & Wolman, 2007), ebenso wie zwischen kognitiven Fähigkeiten und Schulleistung (Chamorro-Premuzic, 2007; Deary, Whiteman, Starr, Whalley & Fox, 2004), ist es von großer Wichtigkeit zu überprüfen, dass die Effekte des FSK auf Schulleistung nicht durch die jeweiligen kognitiven Fähigkeiten konfundiert sind (Chamorro-Premuzic, Harlaar, Greven & Plomin, 2010). Es stellt sich also die Frage nach der relativen Bedeutsamkeit der verschiedenen Faktoren bzw. der inkrementellen Validität motivationaler Faktoren. Schließlich bestreiten einige Autoren, dass motivationale Faktoren auch dann noch in der Lage sind, eigenständige Varianzanteile in der Schulleistung aufzuklären, wenn kognitive Fähigkeiten als erklärungsmächtigste Faktoren gleichzeitig berücksichtigt werden (Gagné & St. Père, 2001). So ergab die Studie von Gagné und St. Père (2001), dass lediglich der IQ, nicht jedoch die selbsteingeschätzte Motivation zur Vorhersage der Schulleistung signifikant beitrug. Allerdings wurde diese Studie an Achtklässlerinnen einer reinen Mädchenschule durchgeführt, so dass von einer selektiven Stichprobe gesprochen werden muss. Die Studie von Schicke und Fagan (1994) sprach hingegen für die inkrementelle Bedeutung von Motivation. Denn in ihrer Untersuchung an Viert-, Sechst- und Achtklässlern erklärte das leistungsbezogene Selbstkonzept über Intelligenz zusätzliche 6.72% im GPA auf. Dennoch erwies sich auch hier Intelligenz als bester Prädiktor. Auch neuere Studien sprechen für die prädiktive Validität motivationaler Faktoren über kognitive Fähigkeiten hinaus (Greven, Harlaar, Kovas, Chamorro-Premuzic & Plomin, 2009). So ergab die Studie von Spinath et al. (2006), dass das fachspezifische FSK 8% an zusätzlicher Varianz in der Mathematikleistung vorhersagen konnte, sowie 9% in Englisch. IM erwies sich hingegen als nicht signifikant über Intelligenz hinaus. In dieser Studie wurden jedoch keine Schulnoten betrachtet, sondern Lehrerratings. Die Studie von Steinmayr und Spinath (2009), in der Schüler der elften und zwölften Klasse untersucht wurden,

ergab ebenfalls, dass FSK sowohl in Deutsch, als auch in Mathematik einen beachtlichen Anteil zusätzlicher Varianz über Intelligenz hinaus aufzuklären im Stande war, nämlich 9% in Deutsch, sowie 30% in Mathematik. FSK erwies sich somit in beiden Fächern gar als stärkster Prädiktor, da verbale Intelligenz für sich genommen lediglich 5% spezifische Varianz in Deutsch erklären konnte und in Mathematik nur 11% durch numerische Intelligenz erklärt wurden. Somit zeigte sich in dieser Studie, ebenso wie in der Studie von Schilling et al. (2006), eine stärkere Assoziation des FSKs zu Mathematik, während Eccles (1984) und Spinath et al. (2006) eine stärkere Assoziation zur sprachlichen Domäne fanden. Ein Grund dieser konträren Befunde könnte sein, dass sich die Studien von Steinmayr und Spinath (2009) sowie von Schilling und Kollegen (2006) auf eine Gymnasialstichprobe bezogen, während die beiden anderen Studien Grundschüler untersuchten. Im Falle der Studie von Eccles (1984) erfolgte die Erfassung des Kriteriums außerdem erst ein Jahr nach der Erfassung des FSKs.

Ein weiterer Unterschied zwischen der Studie von Steinmayr und Spinath (2009) und der Studie von Spinath et al. (2006) ist, dass sich bei Steinmayr und Spinath (2009) die Wert-Komponente, welche in Anlehnung an das Modell von Eccles operationalisiert wurde, ebenfalls als signifikanter Prädiktor erwies. Diese war mit 8% inkrementeller Validität in Deutsch und 26% in Mathematik als ähnlich bedeutsam einzustufen wie FSK. Möglicherweise ist dies als Hinweis auf eine zunehmende Bedeutung von IM mit steigendem Alter zu sehen. Besonders erwähnenswert an der Studie von Steinmayr und Spinath (2009) ist außerdem, dass Noten als Leistungskriterium dienten und dass das Befundmuster sich selbst dann bestätigte, wenn die Vorleistung kontrolliert wurde. Allerdings verringerte die Kontrolle das jeweilige Inkrement.

Weitere wichtige Befunde lieferte die Längsschnittuntersuchung von Chamorro-Premuzic et al. (2010), in der an einer Stichprobe von Neunjährigen der kausale Zusammenhang zwischen FSK und Schulleistung untersucht wurde. Im Gegensatz zu den Studien von Marsh und Kollegen (2005), wurde hier jedoch zusätzlich die kognitive Leistung kontrolliert, um der bereits angesprochenen Interkorrelation zwischen den verschiedenen Faktoren Rechnung zu tragen. Es bestätigte sich nun, dass FSK selbst unter Kontrolle der kognitiven Fähigkeiten die Schulleistung (Lehrerrating) im Alter von zwölf Jahren vorhersagen konnte, ebenso wie die Fähigkeitsüberzeugungen zu diesem Zeitpunkt. Außerdem erwies sich der Einfluss des FSK zum Messzeitpunkt 1 auf die Leistung zu Zeitpunkt 2 als ebenso stark, wie

der Einfluss der Leistung bei der ersten Messung auf das FSK zum Zeitpunkt 2. Damit bestätigte diese Studie zum einen die durch Marsh et al. (2005) postulierten reziproken Zusammenhänge und zum anderen die inkrementelle Validität des FSKs als Prädiktor schulischer Leistung (Chamorro-Premuzic et al., 2010).

Studien, die die inkrementelle Validität von IM nachweisen sind zwar insgesamt seltener. Dennoch existieren neben der bereits genannten Studie von Steinmayr und Spinath (2009) weitere Belege für die inkrementelle Bedeutung von IM über Intelligenz hinaus, sowohl bei Grundschulern (Gottfried, 1990), als auch bei Fünft- bis Achtklässlern (Gottfried, 1985) bzw. Zehntklässlern (Lloyd & Barenblatt, 1998). Zwar lagen allen Studien, mit Ausnahme der zuletzt genannten, kleine Stichproben von $N < 200$ zu Grunde, dennoch konnte eine inkrementelle Validität von 3% bis 6% nachgewiesen werden (Spinath et al., 2006).

Zuletzt bleibt zu erwähnen, dass sich nur wenige Studien bisher mit der Betrachtung des Verhältnisses von spezifischer und geteilter Varianz zwischen Motivation und Intelligenz beschäftigt haben. Dieser Aspekt ist jedoch vor dem Hintergrund der Bewertung der relativen Bedeutsamkeit der verschiedenen Prädiktorgruppen interessant. Denn Gottfried (1990) fand beispielsweise signifikant positive Korrelationen zwischen IM und IQ im Bereich von $r = .25 - .39$ und auch Spinath et al. (2006) berichten schwach positive Zusammenhänge zwischen FSK und g ($r = .18 - .25$). Hinsichtlich des Anteils an gemeinsamer und spezifischer Varianz fand Helmke (1992), dass bei der Vorhersage von Mathematiknoten lediglich 8.9% der erklärten Varianz spezifisch auf Intelligenz zurückzuführen waren. FSK erreichte einen eigenständigen Varianzanteil von 11.0%. Doch der von beiden Prädiktoren geteilte Varianzanteil, welcher 30.4% umfasste, erwies sich als stärkster Faktor. Auch in der Studie von Spinath et al. (2006) konnten 22% geteilte Varianz in Mathematik, sowie 24% in Englisch nachgewiesen werden. Allerdings korrespondierte der stärkste Varianzanteil spezifisch mit Intelligenz (53% in Mathematik, 45% in Englisch). Der Abstand des FSKs zur Intelligenz hinsichtlich des Anteils unguetlicher Varianz erwies sich mit 31% in Englisch als geringer verglichen mit Mathematik, wo lediglich 25% spezifisch mit FSK korrespondierten. Unter Beachtung der bisherigen Befunde nahmen Spinath et al. (2006) an, dass zwischen motivationalen Komponenten und Intelligenz eine wechselseitige Beeinflussung stattfindet, welche auch Auswirkungen auf die Schulleistung haben sollte.

Zusammengefasst sprechen viele Studienergebnisse für die inkrementelle Validität non-kognitiver Prädiktoren wie FSK und IM. Aufgrund der teilweise selektiven

Stichproben (z.B. Gymnasiasten) bzw. der unterschiedlichen Leistungskriterien, ist die relative Bedeutung gegenüber kognitiven Faktoren nicht abschließend geklärt. Außerdem ist unklar, ob die sprachliche oder die mathematische Domäne stärker mit non-kognitiven Faktoren assoziiert ist. Eine gleichzeitige Betrachtung von WM und motivationalen Faktoren im Hinblick auf die relative Bedeutung als Prädiktoren schulischer Leistung wurde bisher versäumt, weshalb die vorliegende Untersuchung die bestehenden Ansätze erweitert.

1.5.3 Sozioökonomischer Status und Schulerfolg

Aufgrund der bereits in Kapitel 1.4 angesprochenen Assoziation von SES und entwicklungsfördernden Umgebungsfaktoren, ist es naheliegend, dass SES sich auch als signifikant im Hinblick auf die Vorhersage des Schulerfolgs erweisen kann. So konnte bereits in verschiedenen Studien ein positiver Zusammenhang nachgewiesen werden (Baumert & Schümer, 2001; Ehmke, Hohensee, Heidemeier & Prenzel, 2004; Gottfried et al., 2003). Auch für spezifische Fähigkeiten konnte eine Abhängigkeit vom familiären SES festgestellt werden (Baumert, Watermann & Schümer, 2003). Beispielsweise ergab die PISA-Studie (OECD, 2001) Zusammenhänge zwischen $r = .25$ und $.41$ zwischen SES und Lese-, Problemlöse- und Mathematikkompetenz (Watermann & Baumert, 2006).

Unter Beachtung des starken Zusammenhangs von SES und Intelligenz stellt sich jedoch die Frage, ob SES auch dann noch zur Vorhersage beitragen kann, wenn zuvor die kognitiven Fähigkeiten kontrolliert wurden. Während in einer deutschen Längsschnittstudie im Sekundarschulalter kein Einfluss von Indikatoren der sozialen Herkunft auf die Leistungsentwicklung von der siebten bis zur zehnten Klasse festzustellen war, nachdem die kognitiven Fähigkeiten und das Vorwissen kontrolliert wurden (Baumert & Köller, 1998), sprechen andere Studien für die inkrementelle Validität des SES (z.B. Schick et al., 2006). So fand auch Kemp (1955) eine Korrelation von $r = .56$ zwischen SES und Schulleistung. Unter Kontrolle der Intelligenz verringerte sich diese jedoch auf einen Wert von $r = .30$, weshalb die Bedeutung des SES verglichen mit Intelligenz als geringer bewertet wurde. Auch White (1982) lieferte hierzu konsistente Befunde, da zwischen SES und Schulleistung lediglich eine Korrelation von $r = .22$ nachgewiesen wurde. Die Studie von Colom und Flores-Mendoza (2007) ergab ebenfalls, dass der Einfluss des SES auf die

Schulleistung zu vernachlässigen ist, besonders im Vergleich zur prädiktiven Validität kognitiver Faktoren. Dennoch wurde im Zuge der PISA-Studie beim Vergleich von Kindern aus Familien mit hohem und niedrigem Bildungsgrad ein Wissensunterschied bezüglich mathematischer Fähigkeiten von mehr als zwei Schuljahren festgestellt (Ehmke, Siegle & Hohensee, 2005). Außerdem wurde eine Varianzaufklärung von 22.3% berichtet, die der SES in der mathematischen Kompetenz von Fünfzehnjährigen aufklären konnte (Ehmke et al., 2005). Nach Baumert, Watermann und Schümer (2003) ist jedoch zu beachten, dass der Einfluss des SES zu einem substantiellen Anteil durch familiäre Prozessmerkmale, wie etwa die Kommunikationspraxis, vermittelt wird.

Baumert und Schümer (2001) warnen vor diesem Hintergrund vor einer vorschnellen Interpretation sozialer Disparitäten im Bildungssystem als Indiz der diskriminierenden Wirkung der Institution Schule, im Sinne einer generellen Benachteiligung von Kindern niedriger sozialer Schichten. Schließlich seien soziologische Disparitätsuntersuchungen überwiegend Umfragestudien, die korrelative Zusammenhänge untersuchen und somit „Ergebnisse von Bildungsentscheidungen und Bildungsverläufen, aber keine Bildungsprozesse“ (Baumert & Schümer, 2001, S. 353) abbilden. Die Autoren gehen weiterhin davon aus, dass sich Benachteiligungen am ehesten am Ende der Grundschulzeit nachweisen lassen, wenn Lehrkräfte Empfehlungen für weiterführende Schulen aussprechen. Denn es existieren einige Hinweise darauf, dass Kinder aus unteren sozialen Schichten bei gleich guten Zensuren seltener eine Gymnasialempfehlung erhalten, als Kinder aus besser gestellten Familien (Ditton, 1992; Lehmann, Peek & Gänsfuß, 1997). Dass in diesem Zusammenhang auch die Einstellung der Eltern selbst von immenser Bedeutung ist, zeigte die IGLU-Studie (2006). Hier ergab sich, dass Eltern mit niedrigerem SES ihre Kinder erst dann als für den Besuch eines Gymnasiums geeignet einschätzten, wenn diese durchschnittlich 100 Punkte mehr in einem Leistungstest erzielten, als Akademikerkinder (Bos et al., 2008).

Auch der Zugang zu bildungsfördernden Medien spielt eine Rolle für die Entwicklung schulischer Fähigkeiten. So kann akademisches Interesse, wie bereits erwähnt, als Prädiktor schulischer Leistung gesehen werden (Eccles & Wigfield, 2002). Akademisches Interesse bildet sich nun mit größerer Wahrscheinlichkeit in einem bildungsaffinen Umfeld aus, welches wiederum Assoziationen zum SES aufweist.

Insgesamt gilt, dass die bisherigen Forschungsergebnisse bezüglich der Bedeutung des SES als Prädiktor schulischer Leistung aufgrund sehr unterschiedlicher

Operationalisierungen von Leistung und SES nur schwer vergleichbar sind, weshalb die prädiktive Validität des SES nicht vollständig geklärt ist. Aus diesem Grund zielt die vorliegende Untersuchung auch auf die Betrachtung der Vorhersagekraft des SES ab.

Im Anschluss an die Darstellung des theoretischen Hintergrunds der Studie erfolgt nun die Herleitung der Fragestellungen und Hypothesen dieser Arbeit.

II Fragestellung und Hypothesen

Die vorliegende Studie fokussiert auf die Untersuchung des relativen Beitrags kognitiver und non-kognitiver Prädiktoren bei der Vorhersage des Schulerfolgs. Auch wenn die Untersuchung der prädiktiven Bedeutung verschiedener individueller Faktoren auf die schulische Leistung bereits Gegenstand zahlreicher Studien war (vgl. Kapitel 1.5), ist das komplexe Zusammenspiel der verschiedenen Determinanten nur unzureichend geklärt.

Da sich viele Studien außerdem auf einzelne Kategorien individueller Determinanten beschränken, wie beispielsweise die Gruppe der kognitive Fähigkeiten, wurde in der vorliegenden Studie ein interdisziplinärer Ansatz verfolgt, indem gleichzeitig Variablen betrachtet wurden, welche typischerweise im Bereich der pädagogischen Psychologie Berücksichtigung finden (FSK, IM), sowie kognitive Variablen aus dem Bereich der differentiellen und kognitiven Psychologie (WM, gf). Durch die zusätzliche Beachtung des familiären SES wurde außerdem eine viel diskutierte Umweltvariable in die Analysen integriert. Besonders die gleichzeitige Betrachtung von WM und gf im Kontext motivationaler Variablen stellt einen neuen Ansatz dar, der durch die Erweiterung bisheriger Perspektiven zur Gewinnung neuer Erkenntnisse beitragen soll.

Basierend auf den bereits dargelegten theoretischen Grundlagen und empirischen Evidenzen werden nun die der Studie zugrunde liegenden Hypothesen hergeleitet.

Während sich der erste Teil der Arbeit mit der vergleichenden Betrachtung der kognitiven Prädiktoren WM und gf beschäftigt, wird im zweiten Teil die relative Bedeutung kognitiver und non-kognitiver Variablen untersucht. In beiden Teilen wird zusätzlich der elterliche SES berücksichtigt. Von besonderem Interesse ist der Aspekt der inkrementellen Validität, also die Frage, inwiefern non-kognitive Variablen im Stande sind, über kognitive Fähigkeiten hinaus zusätzliche Varianz aufzuklären. Alle Analysen werden getrennt für Deutsch und Mathematik durchgeführt, um möglicherweise auftretende fachspezifische Effekte aufdecken zu können.

2.1 Die Bedeutung der kognitiven Prädiktoren WM und gf für die Vorhersage des Schulerfolgs

Kognitive Fähigkeiten gelten mit einer durchschnittlichen Varianzaufklärung von 25% als wichtigste Determinanten des Schulerfolgs (Neisser et al., 1996). So wird auch in der vorliegenden Studie ein positiver Zusammenhang zwischen WM, gf und Schulerfolg erwartet. Ein wichtiges Ziel dieser Studie ist jedoch die vergleichende Betrachtung der beiden kognitiven Faktoren WM und gf hinsichtlich ihrer prädiktiven Validität. Zwar wurde die prädiktive Validität beider Faktoren für sich genommen bereits empirisch nachgewiesen (z.B. Alloway et al., 2005; Kuncel et al., 2004; siehe Kapitel 1.5.1). Ebenso erwiesen ist die Tatsache, dass WM und gf deutliche Anteile an gemeinsamer Varianz aufweisen (z.B. Colom et al., 2003; Engle et al., 1999). Doch inkonsistent sind die Befunde bezüglich der inkrementellen Validität von WM über gf hinaus bzw. von gf über WM hinaus (siehe Kapitel 1.5.1). Der Nachweis eines signifikanten Inkrements würde für eine getrennte und möglichst breite Erfassung beider Konstrukte sprechen, wodurch ein besseres Vorhersageergebnis zu erzielen wäre. Ein Mangel an inkrementeller Validität würde jedoch auf eine deutliche Überschneidung beider Konstrukte hindeuten, so dass der Mehrwert der Erfassung beider Faktoren zu hinterfragen wäre.

Es stellt sich also die Frage nach dem Verhältnis von spezifischer und gemeinsamer Varianz bezüglich des Kriteriums Schulerfolg. Besonders im Grundschulalter wurden diesbezügliche Studien eher selten durchgeführt. Außerdem erfolgte die Operationalisierung des Kriteriums Schulerfolg meist nicht in Form von Schulnoten, so dass die vorliegende Studie einen wichtigen Beitrag zur Ergänzung bisheriger Studienergebnisse leisten kann.

Da sich in bisherigen Studien außerdem Hinweise auf eine höhere Assoziation der kognitiven Faktoren WM und gf zur mathematischen Domäne ergaben (Krumm et al., 2008), wird der Aspekt fachspezifischer Unterschiede auch in der vorliegenden Studie aufgegriffen.

Basierend auf bisherigen Befunden werden folgende Hypothesen bezüglich der prädiktiven Validität von WM und gf spezifiziert.

Hypothese 1a: Gemeinsame Varianz von WM und gf

WM und gf weisen einen substantiellen Anteil an gemeinsamer Varianz auf, der zur Vorhersage des Schulerfolgs signifikant beiträgt.

Hypothese 1b: Spezifische Varianz von WM und gf

WM und gf stellen distinkte Variablen dar, die signifikant zur Prädiktion des Schulerfolgs beitragen. Beide Faktoren klären spezifische Varianzanteile in vergleichbarer Höhe im Kriterium auf, weshalb die beste Prädiktion durch die gleichzeitige Berücksichtigung von WM und gf erzielt wird.

Hypothese 1c: Fachspezifische Effekte

Die prädiktive Validität beider Faktoren kann sowohl in Deutsch als auch in Mathematik nachgewiesen werden, wobei die Vorhersagekraft im Fach Mathematik höher ausfällt als in Deutsch.

2.2 Die relative Bedeutung kognitiver und non-kognitiven Prädiktoren bei der Vorhersage des Schulerfolgs

Trotz des starken Einflusses kognitiver Faktoren bleibt ein beachtlicher Anteil an Varianz im Kriterium Schulerfolg unaufgeklärt. Daher betonen einige Autoren den nicht zu vernachlässigenden Einfluss nicht-fähigkeitsbezogener Eigenschaften, wie beispielsweise der Motivation (z.B. Spinath & Steinmayr, 2009) und verweisen auf die praktische Bedeutung dieser Variable. Denn diese gilt im Allgemeinen als veränderlicher, verglichen mit kognitiven Fähigkeiten, und damit als zugänglicher für Interventionen. Der Einfluss motivationaler Variablen liegt insbesondere bei solchen Schülern nahe, die trotz durchschnittlicher kognitiver Fähigkeiten überdurchschnittliche schulische Leistungen erzielen bzw. trotz ausreichend hoher kognitiver Fähigkeiten hinter den erwarteten Leistungen zurückbleiben.

Verschiedene Studien lieferten bereits empirische Bestätigung für die Bedeutung motivationaler Faktoren im Hinblick auf die Vorhersage schulischer Leistungen (z.B. Fraser et al., 1987; Guay et al., 2003, siehe Kapitel 1.5.2). Vor diesem Hintergrund werden auch in der vorliegenden Studie positive Zusammenhänge zwischen den Faktoren FSK und IM und den Schulnoten in Deutsch und Mathematik erwartet.

Da jedoch der Nachweis von Einzelzusammenhängen unzureichend ist, wenn es darum geht, das komplexe Bedingungsgefüge zu beleuchten, welches den Schulerfolg eines Kindes determiniert, liegt der Fokus der Studie auf der Untersuchung der inkrementellen Validität der non-kognitiven Faktoren (FSK, IM) über die kognitiven Faktoren WM und gf hinaus. Die diesbezüglichen Befunde erweisen sich bisher als nicht konsistent (vgl. Kapitel 1.5.2), weshalb weitere Studien zur Klärung der Fragestellung notwendig sind. Die gleichzeitige Analyse des Einflusses von WM, gf, FSK und IM innerhalb eines Datensatzes stellt dabei eine Erweiterung bisheriger Ansätze dar, der zur Klärung der Relevanz motivationaler Faktoren über kognitive Fähigkeiten hinaus beitragen kann.

Im Folgenden werden die diesbezüglichen Hypothesen spezifiziert.

Hypothese 2a: Inkrementelle Validität der non-kognitiven Prädiktoren

Die non-kognitiven Faktoren FSK und IM stellen wichtige Determinanten des Schulerfolgs dar und leisten über die kognitiven Fähigkeiten hinaus einen inkrementellen Beitrag zur Varianzaufklärung.

Hypothese 2b: Relative Bedeutung der kognitiven und non-kognitiven Prädiktoren

Der Anteil an Kriteriumsvarianz, der durch FSK und IM aufgeklärt werden kann, ist geringer als die durch WM und gf aufgeklärte Varianz.

Hypothese 2c: Relative Bedeutung von FSK und IM

Die Variable FSK weist eine höhere prädiktive Validität auf als die Variable IM.

Hypothese 2d: Fachspezifische Effekte

Die non-kognitiven Faktoren FSK und IM weisen im Fach Deutsch eine höhere prädiktive Validität auf als im Fach Mathematik.

2.3 Die Bedeutung des SES für die Vorhersage des Schulerfolgs

Die Frage nach der prädiktiven Bedeutung des elterlichen SES für den Schulerfolg wurde bereits vielfach diskutiert. Während in der empirischen Bildungsforschung häufig die Position eingenommen wird, dass der SES einen direkten Einfluss auf die Schulleistung von Kindern ausübt (vgl. Baumert & Schümer, 2001; siehe Kapitel 1.5.3), vertreten Forscher aus dem Bereich der kognitiven Psychologie bzw. der Verhaltensgenetik die Auffassung, dass der SES die schulische Leistung nur in einem zu vernachlässigenden Ausmaß beeinflusst bzw. dass der Einfluss durch die enge Assoziation mit der Intelligenz der Eltern, welche sich wiederum an die Kinder weiter vererbt, verursacht wird (z.B. Colom & Flores-Mendoza, 2007).

Da jedoch auch Befunde existieren, die für die inkrementelle Validität des SES über Intelligenz hinaus sprechen (Schick et al., 2006), wird auch in der vorliegenden Studie der elterliche SES in die Analysen einbezogen.

Die diesbezüglichen Hypothesen werden wie folgt spezifiziert.

Hypothese 3a: Zusammenhang zwischen SES und den kognitiven Fähigkeiten

Der elterliche sozioökonomische Status weist Korrelationen zu den kognitiven Fähigkeiten WM und gf auf, wobei die Korrelation zwischen gf und SES höher ausfällt, als zwischen WM und SES.

Hypothese 3b: Inkrementelle Validität des SES

Der sozioökonomische Status kann über die kognitiven und non-kognitiven Faktoren hinaus einen signifikanten Beitrag zur Vorhersage des Schulerfolgs leisten, welcher jedoch niedriger ausfällt als der Beitrag der kognitiven bzw. non-kognitiven Prädiktoren.

III Material und Methoden

3.1 Graduiertenkolleg IRTG

Die vorliegende Studie war Teilprojekt des im April 2008 an der Universität des Saarlandes gestarteten internationalen Graduiertenkollegs International Research Training Group (IRTG) *Adaptive Minds: Neural and Environmental Constraints on Learning and Memory*, welches durch die Deutsche Forschungsgesellschaft (DFG) gefördert wird. Insgesamt umfasste das Graduiertenkolleg neun Themengebiete, innerhalb derer vierzehn Doktoranden eigenständige Projekte bearbeiteten. Eine Besonderheit des IRTG ist, dass es sich um ein deutsch-chinesisches Kooperationsprojekt handelt, in dessen Rahmen eine enge Zusammenarbeit mit dem psychologischen Institut der Chinese Academy of Sciences, Beijing stattfindet. Die Kooperation ermöglicht sowohl deutschen als auch chinesischen Doktoranden, im Rahmen eines mehrmonatigen Austausches, Erhebungen am jeweiligen Partnerinstitut zu realisieren. So konnte für das hier dargestellte Projekt durch die Zusammenarbeit mit der chinesischen Partnerdoktorandin Liping Lu ein gemeinsamer deutsch-chinesischer Datensatz generiert werden. Da jedoch der Vergleich zwischen beiden Stichproben nicht Gegenstand der vorliegenden Dissertation darstellt, wird im Folgenden nicht näher auf die Zusammensetzung der chinesischen Stichprobe bzw. den Ablauf der Testung in China eingegangen⁴.

Neben zahlreichen Lehrstühlen der Fachrichtung Psychologie der Universität des Saarlandes partizipiert auch die Abteilung Neuroradiologie der Universitätsklinik Homburg am IRTG. Ziel des IRTG ist es, zu einem besseren Verständnis von Lern- und Gedächtnisprozessen beizutragen. Es sollen sowohl neuronale Grundlagen dieser Prozesse als auch umweltbedingte Einflüsse beleuchtet werden, innerhalb und zwischen den Kulturen.

Die vorliegende Dissertation war eingebettet in das innerhalb des IRTGs unter Topic F geführten Projekts, welches am Lehrstuhl für Differentielle Psychologie und Psychologische Diagnostik der Universität des Saarlandes angesiedelt ist.

⁴ Informationen diesbezüglich können Lu, Weber, Spinath und Shi (2011) entnommen werden.

3.2 Planung und Durchführung der Datenerhebung

Die Datenerhebung fand im Zeitraum von Januar 2009 bis März 2009 an acht verschiedenen saarländischen Grundschulen statt (siehe Tabelle 1). Es wurde darauf geachtet, dass alle saarländischen Landkreise in der Stichprobe vertreten waren, wenngleich eine Überrepräsentation städtischer Grundschulen nicht vermieden werden konnte. Eine Nacherhebung an zwei weiteren Grundschulen wurde im Juni 2009 durchgeführt. Es wurden ausschließlich Kinder der Klassenstufe 4 untersucht. Die Datenerhebung wurde im Vorfeld vom Ministerium für Bildung, Familie, Frauen und Kultur nach detaillierter Beschreibung des Untersuchungsvorhabens sowie Vorlage aller Untersuchungsmaterialien genehmigt. Die Forderungen nach § 2 der „Verordnung über die Durchführung von Erhebungen zum Zwecke wissenschaftlicher Forschung in Schulen“ wurden bei der Datenerhebung realisiert und berücksichtigt.

Nach Genehmigung der Studie durch das Ministerium wurde zunächst telefonisch Kontakt zu verschiedenen zufällig ausgewählten Schulen hergestellt. Nach kurzer Vorab-Information bezüglich der geplanten empirischen Untersuchung wurden bei Zustimmung durch den Direktor die Lehrer der vierten Klasse persönlich besucht und anhand eines ausführlichen Informationsschreibens über Ablauf und Inhalt der Erhebung in Kenntnis gesetzt. Etwa vierzehn Tage vor der jeweiligen Datenerhebung wurden Elterninformationsschreiben an die Schulen versandt, welche von den Lehrern an alle Viertklässler verteilt wurden. Diese Informationsschreiben enthielten neben Hinweisen zu Inhalt und Ablauf der Datenerhebung außerdem eine Einverständniserklärung, welche von den Eltern unterschrieben vor der Datenerhebung dem Lehrer übergeben werden musste, um die Teilnahme des Kindes an der Studie zu erlauben. Es wurden ausschließlich diejenigen Kinder in die Untersuchung einbezogen, deren Eltern vor dem Erhebungstermin unterschriebene Einverständniserklärungen vorlegten. Den Eltern wurde verdeutlicht, dass es sich um eine freiwillige Teilnahme handelte und dass alle Auflagen des Ministeriums bezüglich Datenschutz und Anonymität umgesetzt wurden.

Die Datenerhebung fand in zwei separaten Phasen statt, einer Schultestung und einer Heimtestung. Die neunzigminütige Schultestung diente der Erfassung der kognitiven Konstrukte Arbeitsgedächtnis und Intelligenz. Die als Gruppentestung konzipierte Erhebung fand klassenweise statt, wobei bei geringer Beteiligung

mehrere Klassen zusammengefasst getestet wurden. Es wurde darauf geachtet, dass die Gruppengröße unterhalb einer Grenze von 25 Kindern blieb. Im Anschluss an die kognitive Testung erhielten die Kinder Umschläge, welche jeweils einen Eltern- und einen Kinderfragebogen zu nicht-kognitiven Einflussfaktoren beinhalteten. Neben den für diese Arbeit relevanten motivationalen Konstrukten wurden zusätzlich in beiden Fragebögen *elterliches Involvement* und *Erziehungsverhalten* erfasst. Der Elternbogen enthielt darüber hinaus noch Fragen bezüglich des sozioökonomischen Status der Familie, einen Fragebogen, der das *kindliche Temperament* erfasste und Fragen zu elterlichen *Normen und Werten*. Da der Fokus dieser Arbeit jedoch auf den individuellen Determinanten des Schulerfolgs lag, fanden neben den Angaben zum sozioökonomischen Status nur die Selbsteinschätzungen des Kindes bezüglich motivationaler Komponenten in den folgenden Analysen Berücksichtigung. Eine Übersicht über alle im Eltern- und Kinderfragebogen enthaltenen Maße, sowie der jeweiligen Erhebungsperspektive (Selbst- vs. Fremdbbericht) findet sich im Anhang A. Die Fragebögen wurden von den Eltern und Kindern zu Hause ausgefüllt und in verschlossenen Umschlägen wieder dem Lehrer übergeben.

Die Erhebung der Schulnoten erfolgte mit Hilfe von Tabellen, welche Codenummer sowie Name jedes Kindes erfassten (siehe Anhang B). Die Codenummern wurden den Kindern vor der kognitiven Testung zugewiesen und sowohl auf den Antwortbögen, als auch auf den Fragebögen vermerkt. Die Anonymität wurde gewährleistet, indem die Lehrer die Namensspalte nach dem Ausfüllen der für die Deutsch- und Mathematiknoten der dritten und vierten Klassenstufe vorgesehenen Spalten der Notentabelle abtrennten, so dass lediglich eine Codenummer für jedes Kind vorhanden war, die keine Zuordnung mehr zu einem einzelnen Schüler erlaubte. Die Notentabellen wurden zusammen mit den Fragebögen acht Tage nach der Testung an der jeweiligen Schule abgeholt.

3.3 Beschreibung der Stichprobe

Insgesamt konnten zehn Schulen für die vorliegende empirische Untersuchung gewonnen werden. Aus der Grundmenge von $N = 634$ Kindern aus 29 Klassen nahmen 320 an der Untersuchung teil. Es stimmten also rund 50% der Eltern der Testung zu. Von 320 kognitiv getesteten Kindern füllten 294 Kinder den Kinderfragebogen aus (92%) und 291 Eltern den Elternfragebogen (90%). Diese

Rücklaufquote kann als äußerst zufriedenstellend bezeichnet werden. Somit lagen $N = 291$ vollständige Eltern-Datensätze vor, sowie $N = 294$ vollständige Kinderdatensätze. Allerdings beziehen sich die Analysen dieser Arbeit auf die Angaben im Kinderfragebogen. Lediglich der höchste Bildungsabschluss der Eltern als Indikator des sozioökonomischen Status, sowie die Informationen bezüglich Alter und Geschlecht des Kindes wurden dem Elternfragebogen entnommen.

Tabelle 1 gibt einen Überblick über die an der Untersuchung beteiligten Schulen, sowie der jeweiligen Schülerzahl pro Schule.

Tabelle 1: Übersicht der an der Studie beteiligten Schulen, der entsprechender Schülerzahl und der Rücklaufquote

Landkreis	Ort	Schülerzahl		Rücklauf FB	
		gesamt	teilnehmend	Kinder	Eltern
Neunkirchen	Schiffweiler	56	34	33	33
	Landsweiler-Reden	50	18	16	16
	Wiebelskirchen	75	17	15	14
	Hüttigweiler	58	38	38	38
St. Wendel	Oberkirchen	60	21	19	18
Saarpfalz-Kreis	St. Ingbert	58	39	33	33
	Homburg	91	65	57	57
Merzig-Wadern	Mettlach	47	19	19	19
Stadtverband Saarbrücken	Sulzbach	64	30	27	27
Saarlouis	Völklingen	75	39	37	36
Gesamt		634	320	294	291

Anmerkung: gesamt: Anzahl Viertklässler an der jeweiligen Schule; teilnehmend: Anzahl der an der Studie teilnehmenden Viertklässler; Rücklauf FB Kinder: Anzahl der bearbeiteten Kinderfragebögen; Rücklauf FB Eltern: Anzahl der bearbeiteten Elternfragebögen.

Die Geschlechterverteilung konnte mit 48.8% weiblichen ($N = 156$) und 41.3% männlichen Schülern ($N = 132$) als annähernd gleichverteilt angesehen werden. Von 10% der Schüler ($N = 32$) fehlten Angaben zum Geschlecht.

Das durchschnittliche Alter der Kinder innerhalb der Gesamtstichprobe (vgl. Tabelle 2) lag bei 9.7 Jahren ($SD = 0.56$), wobei bei 42 Kindern diesbezügliche Angaben fehlten. Innerhalb der Substichprobe von $N = 278$ Kindern, für welche Altersangaben

vorlagen, befanden sich 95.6% der Kinder im für die vierte Klassenstufe zu erwartenden Alter von neun bis zehn Jahren.

Tabelle 2: Altersstruktur und Geschlechterverteilung der Stichprobe

Alter	N	%	Gültige %	Geschlecht	
				m	w
8	1	0.3	0.4	1	-
9	86	26.9	30.9	35	51
10	180	56.3	64.8	84	95
11	9	2.8	3.2	6	3
12	2	0.6	0.7	2	-
Fehlend	42	13.1			
Gesamt	320	100			

Anmerkung: Bei einem Kind im Alter von zehn Jahren fehlen Angaben zum Geschlecht.

Zwei Kinder befanden sich mit einem Alter von zwölf Jahren mehr als drei Standardabweichungen über dem Mittelwert. Daher wurden diese beiden Testpersonen in den nachfolgenden Analysen ausgeschlossen.

Das durchschnittliche Alter der Mütter lag zum Zeitpunkt der Erhebung bei 39.3 Jahren (SD = 4.7), das der Väter bei 42.2 Jahren (SD = 6.0). 76.3% der Kinder (N = 222) lebten mit beiden biologischen Eltern in einem Haushalt, 8.6% (N = 25) bei Stieffamilien. Bei 13.7% der Haushalte (N = 40) handelte es sich um allein erziehende Eltern. Je zwei Kinder stammten aus Adoptiv- bzw. Pflegefamilien.

Im Anschluss erfolgt die Betrachtung des höchsten Schulabschlusses der Eltern. Diese Information kann als Indikator des sozioökonomischen Status der Familien betrachtet werden (Ensminger & Fothergill, 2003) und liefert wertvolle Hinweise hinsichtlich der Repräsentativität der Stichprobe. Tabelle 3 gibt einen Überblick über die Verteilung der verschiedenen Schulabschlüsse in der Stichprobe, getrennt für Mütter und Väter. Außerdem wird der in der jeweiligen Familie vorliegende höchste Schulabschluss wiedergegeben (siehe *Höchster Abschluss*).

Tabelle 3: Höchster Schulabschluss der Eltern in der Stichprobe im Vergleich zu Angaben des Datenreports 2011 (DESTATIS, 2011)

Schulabschluss	Väter		Mütter		Höchster Abschluss		Report DESTATIS	
	N	%	N	%	N	% (g)	N	% (g)
Ohne Abschluss	7	2.2	4	1.3	3	.9 (1)	121	4(4)
Hauptschulabschluss	89	27.8	77	24.1	59	18.3(20.5)	609	20.2(20.3)
Realschulabschluss	77	24.1	117	36.4	103	32.2(35.8)	1066	35.4(35.7)
Fachabitur/Abitur	88	27.5	85	26.6	123	38.4(42.7)	1198	39.8(40.0)
Ohne Angabe	59	18.4	37	11.6	32	10 (-)	18	.6(-)
Gesamt	320	100	320	100	320	100	3012	100

Anmerkungen: Höchster Abschluss bezieht sich auf das Elternteil mit dem höchsten Schulabschluss innerhalb eines Elternpaares. Werte in Klammern beziehen sich auf gültige Prozente (g). Report DESTATIS: Vergleichsdatensatz des Datenreports Deutschland 2011 des Statistischen Bundesamts (DESTATIS). Die Werte geben den höchsten Abschluss eines Elternpaares an⁵.

Wie die Tabelle zeigt, dominierten die Väter in den unteren Kategorien. So waren in der Stichprobe mehr Männer als Frauen ohne Abschluss oder besaßen lediglich einen Hauptschulabschluss. Allerdings dominierten die Väter auch die Kategorie *Fachabitur/Abitur*, was bedeutet, dass mehr Väter als Mütter den höchstmöglichen Schulabschluss erreichten. Die Mütter gaben am häufigsten an, die mittlere Reife erworben zu haben. In dieser Kategorie waren die Mütter stärker repräsentiert als die Väter. Bezüglich der Fälle ohne Angaben ist festzustellen, dass die Ausfallrate der Väter mit 18.4% höher ausfiel als die der Mütter (11.6%). Die Betrachtung des höchsten Schulabschlusses in der Familie ergab, dass in 38.4% der Fälle mindestens ein Elternteil Fachabitur oder Abitur vorweisen konnte und in 32.2% der Realschulabschluss erreicht wurde. Die Subgruppe der Familien mit Hauptschulabschluss als höchstem Abschluss umfasste 18.3%. Lediglich in 0.9% der Fälle wiesen beide Elternteile keinen Schulabschluss auf.

Die Verteilung der Schulabschlüsse der Väter und Mütter wird außerdem in Abbildung 3 dargestellt. Hier werden die jeweils gültigen Prozente abgebildet, wodurch die Dominanz der Männer in den Kategorien Fachabitur/Abitur bzw. Hauptschulabschluss illustriert wird.

⁵ Die Werte der Spalte Report DESTATIS sind Tab. 4a des Datenreports Deutschland 2011 entnommen. Die Werte für N wurden aus der Gesamtzahl und den jeweiligen Prozentangaben pro Abschluss ermittelt.

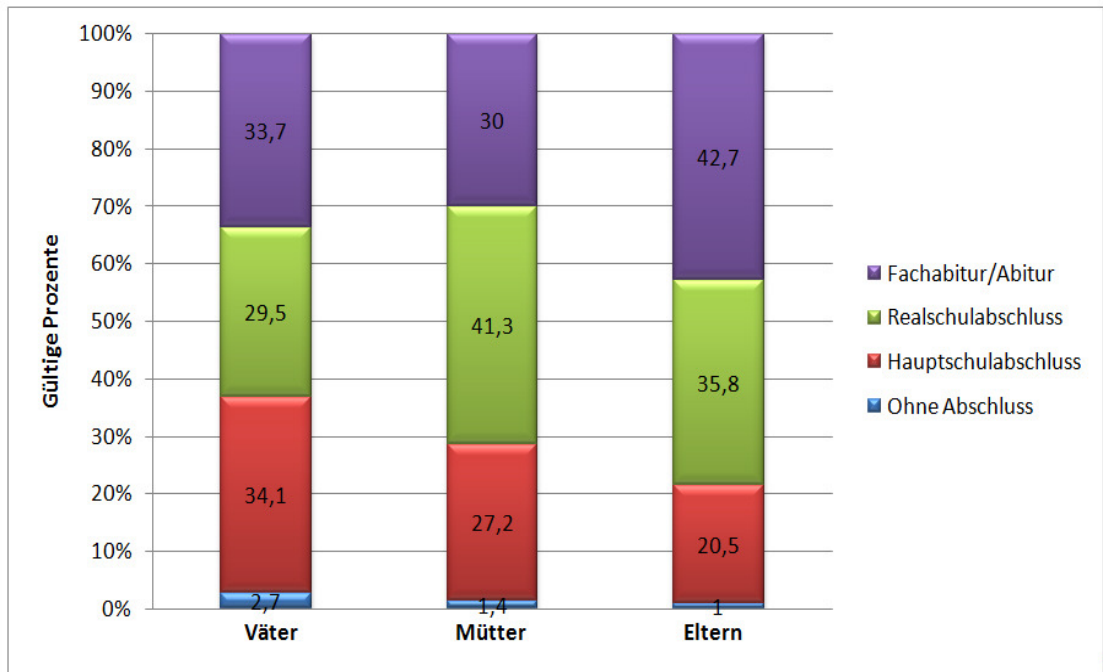


Abbildung 3: Verteilung der Schulabschlüsse innerhalb der Stichprobe

Anmerkungen: N > 261; Werte beziehen sich auf gültige Prozente. Väter: Höchster Schulabschluss des Vaters. Mütter: Höchster Schulabschluss der Mutter; Eltern: Höchster Schulabschluss innerhalb eines Elternpaares.

Die Frage der Repräsentativität der Stichprobe kann jedoch nur im Vergleich zur Verteilung der Schulabschlüsse einer repräsentativen Vergleichsstichprobe beantwortet werden. Dazu wurden dem Datenreport 2011 des statistischen Bundesamtes (DESTATIS) die Zahlen des Erhebungsjahres 2009 entnommen und in Tabelle 3 dargestellt. Diese Referenzwerte beziehen sich auf eine Stichprobe von N = 3012 Personen, deren Kinder im Erhebungsjahr die Grundschule besuchten. Es wurde jeweils der höchste Schulabschluss innerhalb eines Elternpaares berücksichtigt. Zur besseren Vergleichbarkeit wurden in Abbildung 4 die jeweils gültigen Prozente der beiden Stichproben gegenübergestellt. Die Anzahl der Fälle ohne Angaben kann Tabelle 3 entnommen werden⁶.

⁶ In Tabelle 3 bzw. Abbildung 3 und 4 wurde die Kategorie *Studium* mit der Kategorie *Fachabitur/Abitur* subsumiert, um analoge Kategorien zum Datenreport Deutschland (2011) zu erhalten. In späteren Analysen ging die Kategorie *Studium* als höchster Bildungsabschluss ein.

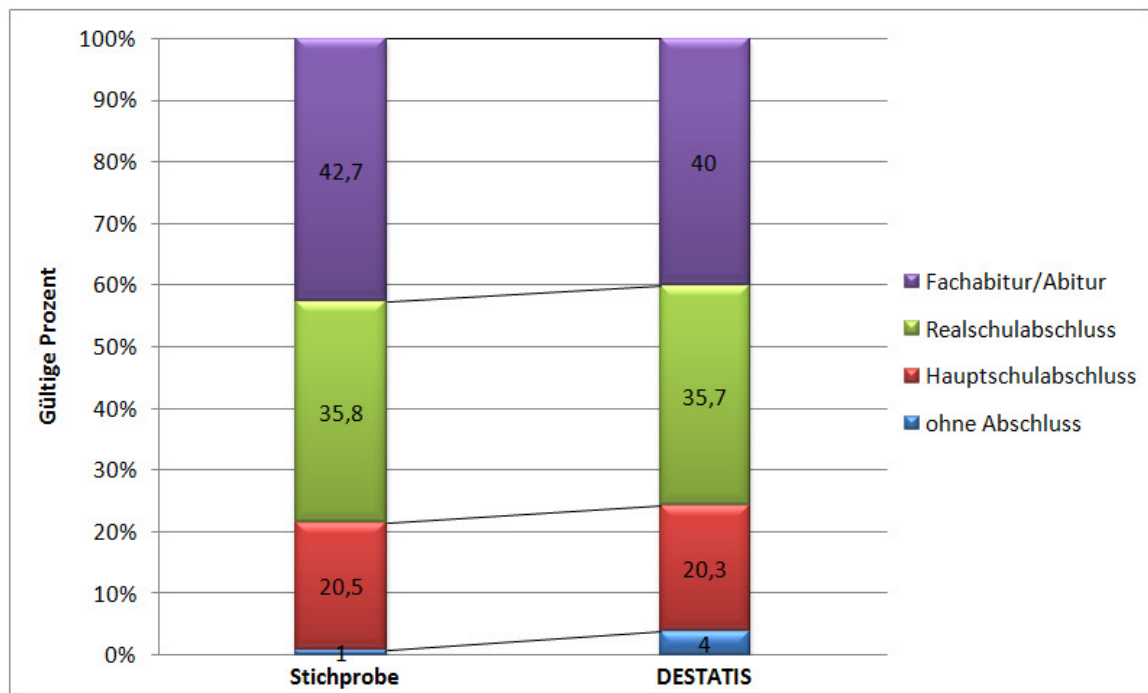


Abbildung 4: Verteilung der Schulabschlüsse der Erhebungsstichprobe im Vergleich zur DESTATIS-Stichprobe

Wie Abbildung 4 zeigt, kann die erhobene saarländische Stichprobe als weitestgehend repräsentativ betrachtet werden. Es fand sich lediglich eine leichte Unterrepräsentation an Personen ohne Schulabschluss im Vergleich zur DESTATIS-Stichprobe. Außerdem lag eine leichte Überrepräsentation von rund 3% an Eltern mit Fachabitur bzw. Abitur vor. Die Kategorien Realschulabschluss und Hauptschulabschluss erwiesen sich als kaum von der Verteilung der Vergleichsstichprobe abweichend. Im Hinblick auf die Anzahl der Fälle ohne Angaben ist festzustellen, dass in der saarländischen Stichprobe ein höherer Prozentsatz an Missings auftrat als in der Vergleichsstichprobe (Missings_{IRTG}: 10%, Missings_{DESTATIS}: 0.6%). Die Repräsentativität der erhobenen Stichprobe kann als zufriedenstellend betrachtet werden.

3.4 Erhebungsinstrumente

3.4.1 Erfassung der kognitiven Fähigkeiten

3.4.1.1 *Arbeitsgedächtnis*

Zur Messung der Arbeitsgedächtniskapazität wurden drei verschiedene Aufgaben eingesetzt, welche Adaptationen bzw. Weiterentwicklungen gängiger WM-Aufgaben darstellen und auf Baddeley's Multi-Komponenten-Modell basieren. Die drei ausgewählten Storage-and-Processing Aufgaben enthielten sowohl verbales als auch visuell-räumliches Stimulusmaterial und beanspruchten nicht nur die Slave-Systeme, sondern auch die ZE, da nicht nur das Abspeichern der Stimuli erfordert wurde, sondern auch die Transformation des gehaltenen Inputs. Um den Schwierigkeitsgrad der Aufgaben so zu wählen, dass Boden- und Deckeneffekte vermieden wurden und um sicherzustellen, dass alle Instruktionen verstanden wurden, die Antwortzeiten ausreichend waren und die vorgegebene maximale Testzeit von zwei Schulstunden eingehalten werden konnte, wurden vor Beginn der Datenerhebung zwei Pilottests durchgeführt, in denen alle Aufgaben von Viertklässlern ($N = 15$) bearbeitet wurden, die nicht der eigentlichen Stichprobe angehörten. Im Anschluss an jede Pilotierung wurden die Aufgaben optimiert. Im Folgenden werden die Versionen beschrieben, die bei der tatsächlichen Datenerhebung zum Einsatz kamen.

Eine Anpassung verbreiteter Aufgaben war deshalb vonnöten, da das Design der vorliegenden Studie eine Gruppentestung im Klassenverband erforderte, um den Verlust von Unterrichtszeit für alle Beteiligten so gering wie möglich zu halten. Da es nicht möglich war, jedem Kind für die Testung einen eigenen PC zur Verfügung zu stellen, mussten außerdem Aufgaben herangezogen werden, welche eine Paper-Pencil-Testung erlaubten, was ebenfalls die Einsetzbarkeit vieler gängiger WM-Tests stark beeinträchtigte. Eine weitere Einschränkung war, dass die Aufgaben keine Verbalisierungselemente beinhalten durften, wie beispielsweise lautes Nachsprechen von Wörtern oder Zahlen, da dies im Gruppenverband zu starken Störeffekten geführt hätte. Auf inhaltlicher Ebene musste die Einsetzbarkeit der Aufgaben in einer deutschen und in einer chinesischen Stichprobe gewährleistet sein. Hierzu wurden bei der Itemgenerierung Indikatoren wie Silbenzahl pro Wort, Bekanntheit bzw. Vertrautheit der Stimuli und Worthäufigkeit in beiden Nationen berücksichtigt.

Unter Beachtung dieser zeitlichen, inhaltlichen und organisatorischen Einschränkungen kamen schließlich die folgenden drei Aufgaben zum Einsatz.

Digit Sorting (DS)

In dieser sowohl die PL als auch die ZE beanspruchenden Aufgabe wurden die Kinder aufgefordert, sich Zahlenreihen zu merken und anschließend sortiert nach aufsteigender Reihenfolge zu reproduzieren. Alle Zahlenreihen wurden per Lautsprecher dargeboten (z.B. 5, 3). Zur Reproduktion der sortierten Listen wurden Antwortblätter bereitgestellt. Die Trials umfassten Ziffern zwischen 1 und 12 (ohne 7⁷), welche mit einer Geschwindigkeit von einer Ziffer pro Sekunde präsentiert wurden. Alle 11 Zahlen wurden vor der Testung ausgesprochen und anschließend zu einzelnen Trials zusammengeschnitten, um Sprechdauer, Tonfall, Deutlichkeit und Lautstärke zu kontrollieren. Insgesamt galt es zehn Trials zu bearbeiten, je zwei mit gleicher Listenlänge, wobei diese zwischen zwei und sechs Ziffern variierte. Listen unterschiedlicher Länge wurden durchmischt dargeboten. Je nach Listenlänge standen bis zu 15 Sekunden Antwortzeit zur Verfügung, danach startete automatisch der nächste Trial. Es wurde darauf geachtet, dass während der Stimuluspräsentation nicht geschrieben wurde. Die Gruppierung der Ziffern zu Trials erfolgte zufällig, wobei auf eine annähernde Gleichverteilung aller 11 Ziffern über alle Trials hinweg geachtet wurde. Ebenso wurde kontrolliert, dass gleiche Ziffern nicht mehrmals innerhalb eines Trials auftraten. Die zehn eingesetzten Zahlenreihen sind in der dargebotenen Reihenfolge Anhang B zu entnehmen.

Vor Beginn der eigentlichen Testung wurden den Kindern alle Zahlen vorgespielt, um sie mit dem Klang des Stimulusmaterials vertraut zu machen. Anschließend wurden drei Trainingstrials durchgeführt und im Klassenverband besprochen, um zu gewährleisten, dass alle Kinder die Instruktion verstanden hatten. Die gesamte Testdauer für diese Aufgabe lag mit vorheriger Instruktion bei ca. 15 Minuten.

Die richtig gelösten Aufgaben wurden zu einem ungewichteten Summenscore verrechnet. Aufgaben, welche entgegen der Instruktion bearbeitet wurden, indem z.B.

⁷ Die Zahl 7 wurde nicht in den Itempool aufgenommen, um ausschließlich einsilbige Ziffern einsetzen zu können.

während der Stimuluspräsentation Notizen am Rand vermerkt wurden, wurden als Missing bewertet.

Animal Sorting (AS)

In dieser Aufgabe wurden Listen mit Tiernamen präsentiert (z.B. Pferd, Maus), welche sich die Kinder merken und sortiert nach aufsteigender Körpergröße auf einem Antwortbogen aufschreiben sollten. Auch diese Listen wurden per Lautsprecher dargeboten. Alle Tiernamen wurden vorab eingesprochen und zu Trials zusammengeschnitten. Insgesamt dienten 11 verschiedene Tiere als Stimulusmaterial (siehe Anhang B). Diese Aufgabe vereinte verbale Aspekte (Merken und Aufschreiben der Tiernamen) mit visuell-räumlichen Aspekten (Vergleich der Körpergrößen) und beanspruchte ebenfalls die ZE, da die Items in umsortierter Weise reproduziert werden mussten. Die Anzahl der verwendeten Tiere wurde auf 11 begrenzt, analog zu 11 verschiedenen Ziffern der DS-Task. Die Auswahl der Items erfolgte anhand verschiedener Kriterien. So wurde zum einen darauf geachtet, dass alle Items hinsichtlich der gängigen Normen *frequency*, *name agreement*, und *familiarity* (Snodgrass & Vanderwart, 1980; Yoon et al., 2004) vergleichbar waren, sowohl in Deutschland als auch in China. Zusätzlich wurde die Silbenzahl im Deutschen und im Chinesischen dahingehend kontrolliert, dass das Verhältnis von ein-, zwei- und dreisilbigen Tiernamen zwischen den Stichproben ausgewogen war. Außerdem wurde sichergestellt, dass ein Itempool generiert wurde, der die Erstellung von Trials mit eindeutiger Lösung ermöglichte. Die zehn dargebotenen Tierlisten sind in Anhang B enthalten.

Auch in dieser Aufgabe galt es zehn Trials zu bearbeiten, mit zwischen zwei bis sechs Items variierender Listenlänge. Listen verschiedener Längen wurden durchmischt dargeboten, wobei insgesamt jede Listenlänge zweimalig auftrat. In Abhängigkeit von der Listenlänge standen bis zu 15 Sekunden zur Reproduktion zur Verfügung, danach startete automatisch der nächste Trial. Es wurde darauf geachtet, dass die Kinder während der Stimuluspräsentation nicht schrieben.

Die Gruppierung der Tiere zu Trials erfolgte zufällig, wobei auf eine annähernde Gleichverteilung aller 11 Tiernamen über alle Trials hinweg geachtet wurde. Ebenso wurde kontrolliert, dass es für jeden Trial eine eindeutige Lösung gab.

Vor der eigentlichen Testung wurden alle Tiere über einen an den Rechner angeschlossenen Beamer visuell dargeboten. Die Kinder wurden aufgefordert, den

jeweiligen Namen des Tieres laut mitzusprechen. So wurde sichergestellt, dass alle Tiere bekannt waren. Gleichzeitig wurden die Kinder mit dem Stimulusmaterial vertraut gemacht. Anschließend erfolgte die Instruktion mittels dreier Trainingstrials, die im Klassenverband besprochen wurden, um zu gewährleisten, dass alle Kinder die Aufgabe verstanden. Die gesamte Testdauer für diese Aufgabe lag mit vorheriger Instruktion bei ca. 15 Minuten.

Die richtigen Lösungen dieses Subtests wurden zu einem ungewichteten Summenscore verrechnet. Trials, die entgegen der Instruktion bearbeitet wurden (z.B. Notizen oder Abkürzungen am Blattrand), wurden als Missing gewertet.

Backward Matrix Span (MS)

In dieser visuell-räumlichen Aufgabe wurden den Kindern Matrizen mit 3 x 3 Feldern präsentiert, in denen nacheinander verschiedene Felder rot aufleuchteten. Die Aufgabe der Kinder bestand darin, sich die Reihenfolge der rot aufleuchtenden Felder zu merken und anschließend in der umgekehrten Reihenfolge, also beginnend mit dem zuletzt präsentierten Feld, zu reproduzieren. Die Präsentation der Trials erfolgte mit Hilfe eines an einen Laptop angeschlossenen Beamers, zur Reproduktion wurden Antwortbögen bereitgestellt. Diese Bögen bestanden aus zehn leeren Matrizen, eine Matrix für jeden Trial. Die Kinder wurden instruiert, mit Hilfe von Ziffern in die leeren Matrizen einzutragen, in welcher Reihenfolge die Quadrate präsentiert wurden, wobei darauf geachtet wurde, dass mit dem letzten Item der Liste begonnen wurde. Da die Vortests zeigten, dass es die Kinder überforderte, nicht nur mit dem letzten Item der Liste zu beginnen, sondern zusätzlich auch die Nummern umzukehren, wurde entschieden, dass die Ziffern identisch zur Vorwärts-Abfolge blieben und dass lediglich beim Eintragen der Ziffern mit der letzten (und damit höchsten) Ziffer begonnen werden sollte. Die Quadrate wurden mit einer Geschwindigkeit von zwei Sekunden pro Quadrat dargeboten. Das Ende eines jeden Trials wurde durch das Erscheinen von drei schwarzen Fragezeichen signalisiert. Insgesamt galt es zehn Trials zu bearbeiten. Die Aufgabenlänge variierte zwischen zwei bis sechs Items, wobei jede Listenlänge durch zwei Trials repräsentiert wurde. Zur Reproduktion standen je nach Itemanzahl bis zu 11 Sekunden zur Verfügung, danach startete automatisch der nächste Trial.

Es wurde darauf geachtet, dass während der Stimuluspräsentation nicht geschrieben wurde.

Die Gruppierung der Quadrate zu Trials erfolgte zufällig, wobei auf eine annähernde Gleichverteilung aller neun Quadrate geachtet wurde. Ebenso wurde kontrolliert, dass kein Trial aus ausschließlich benachbarten Quadraten bestand.

Vor der eigentlichen Testung wurde mit Hilfe einer durch Powerpoint unterstützten Instruktion genau erklärt, in welcher Reihenfolge die Kästchen zu markieren waren. Anschließend wurden drei Trainingstrials durchgeführt und im Klassenverband besprochen, um zu gewährleisten, dass alle Kinder die Aufgabe verstanden. Die gesamte Testdauer für diese Aufgabe lag mit vorheriger Instruktion bei ca. 15 Minuten.

Alle richtigen Lösungen dieses Subtests wurden zu einem ungewichteten Summenscore verrechnet, wobei Trials, die nicht instruktionsgemäß bearbeitet wurden (z.B. Eintragen der Ziffern in der Präsentationsrichtung und nicht rückwärts) wurden als Missings kodiert. Die zehn zu bearbeitenden Trials sind Anhang B zu entnehmen.

3.4.1.2 Intelligenz

Die Erhebung der Intelligenz erfolgte ebenso wie die Testung des Arbeitsgedächtnisses im Klassenverband. Es musste also ein für Gruppentestungen geeignetes Paper-Pencil-Verfahren gewählt werden, dessen Durchführung nicht mehr als 45 Minuten beansprucht, um den Verlust an Unterrichtszeit so gering wie möglich zu halten.

Bei dem eingesetzten Inventar handelte sich um den Grundintelligenztest (CFT- 20-R) nach Weiß (2006), welcher ein ökonomisches Verfahren zur Diagnose der grundlegenden geistigen Leistungsfähigkeit darstellt. Der CFT-20-R erfasst sprachfrei mit Hilfe anschaulicher Darstellungen im Paper-Pencil-Verfahren die fluide Intelligenz nach dem Konzept von Cattell, sprich die Fähigkeit figurale Beziehungen und komplexe Zusammenhänge zu erkennen und zu verarbeiten. Das Verfahren ist im Altersbereich von 8.5 bis 60 Jahren einsetzbar, Reliabilität und Konstruktvalidität sind (verglichen mit g) mit $r = .96$ bzw. $r = .78$ bis $.83$ (Bajard, 1955; zitiert nach Weiß, 2006) als gut bis sehr gut zu bewerten. Ein großer Vorteil des CFTs gegenüber anderen Intelligenztests liegt in der ausschließlichen Verwendung von sprachfreiem Testmaterial, ein Aspekt, der für die vorliegende Untersuchung von großer Relevanz war. So belegen Studien zur externen Validität des CFT-20-R im Vergleich zu

sprachbasierten Verfahren geringere Zusammenhänge zu sowohl sozio-kulturellem Status als auch zu Migrationshintergrund (Weiß, 2006). Hieraus kann auf eine geringere Status- und Kulturabhängigkeit des CFT 20-R gegenüber verbalen Intelligenz- oder Schulleistungstests geschlossen werden (ebd.). Durch die Verwendung sprachfreier Aufgaben werden Kinder mit schlechten Sprachkenntnissen und mangelhaft ausgebildeten Kulturtechniken weniger benachteiligt.

Der CFT-20-R besteht aus zwei Testteilen mit je vier Subtests. Aus zeitlichen Gründen wurde in dieser Studie lediglich Testteil 1 eingesetzt. Ebenso wurde auf den Einsatz der ergänzenden Wortschatz- und Zahlenfolgetests verzichtet. Die Itemzahl der vier eingesetzten Subtests variierte zwischen 11 und 15. Der Test ist als power-speed-Test zu verstehen, da sowohl die Schwierigkeit der Items ansteigt, als auch ein Bearbeitungslimit vorgegeben wird. Alle vier Subtests wurden mit der für Gruppentestungen empfohlenen verlängerten Testzeit von vier bis fünf Minuten durchgeführt. Insgesamt wurden 56 Multiple-Choice-Items bearbeitet, bei einer reinen Testzeit von 18 Minuten. Inklusive Instruktion war der CFT-20-R in ca. 40 Minuten durchzuführen, weshalb er optimal in das Zeitfenster von einer Schulstunde passte.

Subtest 1, bestehend aus 15 Items, erforderte das Fortsetzen von Reihen, wofür fünf Minuten zur Verfügung standen. Aus fünf Antwortmöglichkeiten war diejenige Antwort auszuwählen, die die vorgegebene Reihe sinnvoll ergänzte. Subtest 2 umfasste ebenfalls 15 Items. Dieser Subtest zielte auf das Erkennen von Klassifikationen ab. Aus fünf verschiedenen Items galt es das Item auszuwählen, welches nicht in die Gruppe der anderen Stimuli passte. Das zeitliche Limit lag bei fünf Minuten. Der aus 15 Items bestehende Subtest 3 verlangte das Fortsetzen von Matrizen, innerhalb eines zeitlichen Limits von vier Minuten. Subtest 4 erforderte das Erkennen von Topologien. Aus fünf Antwortmöglichkeiten musste diejenige Antwort erkannt werden, in der sich ein abgebildeter Punkt zu den restlichen geometrischen Merkmalen des Bildes so verhielt, wie in der Vorlage. Subtest 4 bestand aus 11 Items, die Bearbeitungszeit war auf vier Minuten limitiert.

Vor Beginn eines jeden Subtests wurden drei Trainingsaufgaben bearbeitet, an der Tafel aufgemalt und im Klassenverband ausführlich besprochen und erklärt. Ebenso wurde das Umblättern der Aufgabenblätter bis zum Ende des jeweiligen Subtests instruiert und geübt. Weiterhin wurde an einem Beispiel dargestellt, wie Korrekturen vorzunehmen waren. Das Ende der Testzeit wurde jeweils eine Minute vor Ablauf

angekündigt. Anders als im Manual beschrieben, sollten die Kinder die Antworten nicht auf ein separates Antwortblatt übertragen, sondern direkt in der jeweiligen Aufgabe das für richtig befundene Item ankreuzen. Korrekturen konnten vorgenommen werden, indem das fälschlicherweise angekreuzte Item umkreist und das richtige Item angekreuzt wurde. Für jeden Subtest wurde ein ungewichteter Summenscore berechnet. Für jede richtig gelöste Aufgabe wurde ein Punkt vergeben.

3.4.2 Erfassung der non-kognitiven Faktoren

3.4.2.1 Fähigkeitsselbstkonzept

Die Erfassung des Fähigkeitsselbstkonzepts erfolgte dem Modell nach Eccles entsprechend anhand dreier fachspezifischer Items pro Schulfach (Eccles (Parsons) et al., 1983). Die Formulierung der Items war stark an den Inhalt des jeweiligen Schulfachs angelehnt. So bezogen sich die Items für das Fach Deutsch auf die Inhalte Lesen, Aufsatz schreiben und Rechtschreibung. Im Fach Mathematik wurden Kopfrechnen, Textaufgaben und Multiplizieren/Dividieren als typische Unterrichtsinhalte erfasst. Die Aufgabe der Kinder war es, für jedes Item mittels einer fünffach gestuften Likert-Skala einzuschätzen, wie gut sie in der jeweiligen Aktivität sind. Die Skala reichte von *gar nicht gut* (1) bis *sehr gut* (5), so dass höhere Werte mit einem höheren Fähigkeitsselbstkonzept einhergingen. Die Antworten wurden getrennt für Deutsch und Mathematik zu einem ungewichteten fachspezifischen Summenscore verrechnet, wenn mindestens zwei Items pro Fach beantwortet wurden. Alle Items der fachspezifischen Skalen können Anhang B entnommen werden.

3.4.2.2 Intrinsische Motivation

Die Erfassung der intrinsischen Motivation erfolgte analog zum Fähigkeitsselbstkonzept mittels dreier fachspezifischer Items pro Schulfach, die die gleichen Unterrichtsinhalte repräsentierten. Im Unterschied zum FSK wurde von den Kindern eingeschätzt, wie sehr sie die jeweiligen Aktivitäten mögen, ebenfalls anhand einer fünffach abgestuften Likert-Skala, die von *mag es überhaupt nicht* (1) bis *mag*

es *sehr* (5) variierte. Somit entsprachen höhere Werte einer höheren intrinsischen Motivation. Die jeweiligen Antworten wurden getrennt für Deutsch und Mathematik zu einem ungewichteten fachspezifischen Summenscores verrechnet, wenn mindestens zwei Items pro Fach beantwortet wurden. Anhang B beinhaltet alle Items mit dazugehöriger Antwortskala.

3.4.3 Erfassung des sozioökonomischen Status

Als Indikator des sozioökonomischen Status diente der höchste Bildungsabschluss der Eltern, welcher getrennt für Mütter und Väter erfasst wurde. Wie bereits in Kapitel 3.3 dargestellt, wurden die sechs Kategorien *ohne Angabe*, *ohne Abschluss*, *Hauptschulabschluss*, *Real-*, *Handels-*, *Gesamtschule mit Abschluss*, *Fachabitur/Abitur* und *Studium* unterschieden. Die beiden Variablen für Mütter und Väter wurden in den Strukturgleichungsanalysen als Indikatoren eines latenten SES-Faktors eingesetzt. Es wurde keine Gewichtung vorgenommen.

3.4.4 Erfassung der Schulnoten

Die Erfassung der Schulnoten erfolgte mit Hilfe der Lehrer. Diese wurden gebeten Notenformulare auszufüllen, in denen sowohl die Jahresnoten der dritten Klasse abgefragt wurden, als auch die Halbjahresnoten der vierten Klasse, jeweils für Deutsch und Mathematik. Die beiden fachspezifischen Noten wurden als Indikatoren der latenten Kriteriumsvariable herangezogen (siehe Anhang B).

3.5 Analysemethoden

3.5.1 Kommunalitätenanalyse

Ein Verfahren, welches wichtige Informationen liefert, wenn die Untersuchung der relativen Bedeutung mehrerer Prädiktoren im Vordergrund steht, ist die *Kommunalitätenanalyse*. Diese Methode ist eng mit der hierarchischen Regressionsanalyse verbunden, welche darauf abzielt, die inkrementelle Validität eines Prädiktors oder einer Prädiktorgruppe über andere Prädiktoren hinaus aufzudecken und die aufgeklärte Kriteriumsvarianz zu maximieren. Die isolierte Betrachtung des inkrementellen Vorhersagebeitrags überschätzt jedoch die Bedeutung des aufklärungsmächtigsten Prädiktors hinsichtlich der Varianzaufklärung im Kriterium und unterschätzt die prädiktive Validität eines schwächeren Prädiktors. Denn korrelieren zwei Prädiktoren miteinander, so liegen üblicherweise auch konfundierte Varianzanteile vor (Spinath et al., 2006). Dieser Anteil an gemeinsamer Varianz wird im regressionsanalytischen Ansatz jedoch stets dem stärksten Prädiktor zugesprochen oder dem Faktor, der zuerst in die Analyse eingegeben wurde (Amado, 2003; Spinath et al., 2006). Die Kommunalitätenanalyse stellt nun ein Verfahren dar, das insbesondere im Falle korrelierender Prädiktoren geeignet ist, eine Zerlegung der aufgeklärten Kriteriumsvarianz vorzunehmen, um die relative Bedeutung der einzelnen Faktoren zu beleuchten (Cooley & Lohnes, 1976). Diese Methode ist nicht hierarchisch, was bedeutet, dass alle Variablen als gleichwertige Einflussfaktoren behandelt werden. Die Kommunalitätenanalyse erlaubt nun die Unterscheidung zwischen der für einen Prädiktor spezifischen Varianz und der gemeinsamen Varianz, die verschiedene Prädiktoren miteinander teilen und die keinem bestimmten Faktor eindeutig zugewiesen werden kann (Spinath et al., 2006). Zur Kalkulation der spezifischen und gemeinsamen Varianzanteile werden die jeweiligen Partkorrelationen der Prädiktoren benötigt, ebenso wie die insgesamt aufgeklärte Kriteriumsvarianz, weshalb zunächst eine Regressionsanalyse durchgeführt wird. Die quadrierte Partkorrelation zwischen Prädiktor und Kriterium wird nun geteilt durch den insgesamt aufgeklärten Anteil an Kriteriumsvarianz (R^2), wodurch sich der spezifische Varianzanteil des jeweiligen Prädiktors ergibt. Subtrahiert man nun die auf diese Weise errechneten spezifischen Varianzanteile von der insgesamt aufgeklärten Varianz, ergibt sich der gemeinsame Varianzanteil (Cooley & Lohnes, 1976). Üblicherweise erfolgt die Angabe der geteilten und spezifischen Varianzanteile in Prozentwerten, wobei die Anteile spezifischer und

gemeinsamer Varianz an der insgesamt aufgeklärten Varianz relativiert werden. Dabei wird die insgesamt aufgeklärte Kriteriumsvarianz auf 100% gesetzt, so dass sich die spezifischen und gemeinsamen Varianzanteile zu 100% aufaddieren (Cooley & Lohnes, 1976).

3.5.2 Strukturgleichungsmodellierung

Der Begriff der *Strukturgleichungsmodellierung* (structural equation modeling, SEM) bezieht sich nicht auf eine spezifische Analysemethode. Vielmehr subsumiert der Begriff eine ganze Familie verwandter Prozeduren, welche sowohl zur Analyse experimenteller als auch nicht experimenteller Datensätze eingesetzt werden können. Allgemein kann die Methodik der SEM als integrativer konfirmatorischer Ansatz bezeichnet werden, in welchem die exploratorische Faktorenanalyse und die Methode der Pfadanalyse zusammengeführt wurden (Kline, 2005). Zur Durchführung der SEM steht spezielle Analysesoftware zur Verfügung. In der vorliegenden Studie wurde AMOS 19 eingesetzt.

Die SEM ist ein Verfahren, welches a priori die Spezifikation eines zu testenden Modells verlangt, das aus einer bestehenden Theorie oder aus bisherigen Befunden abzuleiten ist (Kline, 2005). „Den Ausgangspunkt der Strukturgleichungsmodellierung bildet deshalb immer die Konzeption einer empirisch prüfbar Theorie“ (Weiber & Mülhhaus, 2010, S. 3). Es wird nun untersucht, welche Beziehung zwischen einer oder mehreren unabhängigen Variablen und einer oder mehreren abhängigen Variablen besteht (Tabachnik & Fidell, 2007). Damit kann die multiple Regression, welche die Beziehung mehrerer beobachteter Prädiktoren zu einem beobachteten Kriterium untersucht, als einfachster Fall der SEM betrachtet werden (Tabachnik & Fidell, 2007).

Zentrales Element der SEM ist die Prüfung der Güte der Modellschätzung. Dabei wird bewertet, wie gut ein spezifiziertes Modell die vorliegende Datenstruktur im Sinne einer beobachteten Kovarianzmatrix widerspiegelt. Ziel ist eine minimale Abweichung zwischen beobachteter und vorhergesagter Kovarianzmatrix. Es gilt stets auch Alternativmodelle zu schätzen, um abzusichern, dass ein anderes Modell keine gleich gute Schätzung liefert (Kline, 2005). Es wird deutlich, dass Kovarianzen von zentraler Bedeutung sind im Kontext der Strukturgleichungsmodellierung. Dennoch gehen die Möglichkeiten der Modellschätzung über die Betrachtung von

Kovarianzen hinaus. So können beispielsweise auch Mittelwerte zur Modellschätzung herangezogen werden (Kline, 2005). Zusammengefasst zielt die SEM also darauf ab, ein Modell zu generieren, welches einerseits theoretisch fundiert ist und andererseits eine angemessene statistische Passung zu den beobachteten Daten aufweist (Kline, 2005). Die Bewertung der Passung des theoretischen Modells zu den empirischen Daten erfolgt mittels so genannter *Fit-Indizes*. Diese werden in Abschnitt 3.5.2.3 besprochen. Generell spiegelt der Modellfit wider, wie sehr die modellimplizierte Kovarianzmatrix mit der empirischen Kovarianzmatrix übereinstimmt (Schermmelleh-Engel, Moosbrugger & Müller, 2003). Neben einer möglichst akkuraten Modellschätzung wird außerdem angestrebt, ein möglichst sparsames Modell zu spezifizieren (Parsimonitätskriterium). Dies bedeutet, dass bei gleich guter Passung dem Modell mit den wenigsten Schätzparametern der Vorzug zu geben ist.

Zentrales Kennzeichen der Strukturgleichungsmodellierung ist, dass neben *manifesten* (beobachtbaren) Variablen auch *latente* (nicht beobachtbare) Variablen geschätzt werden können, wodurch ein höherer Abstraktionslevel erreicht wird als bzw. bei herkömmlichen Regressionsanalysen (Kline, 2005). Die Schätzung latenter Variablen erfolgt, indem mehrere manifeste Variablen als Indikatoren der latenten Variablen herangezogen werden. Üblicherweise wird eine SEM mit latenten Variablen als Kausalanalyse bezeichnet (Weiber & Mülhhaus, 2010). Dadurch, dass die SEM darauf abzielt, ein Gesamtmodell zu bewerten und es sich außerdem um ein Verfahren handelt, welches große Stichproben erfordert, ändert sich die Bedeutung statistischer Signifikanztests im Vergleich zu anderen gängigen Analysemethoden. Während die Frage nach der statistischen Signifikanz üblicherweise zentrales Element der Datenanalyse darstellt, verliert dieser Aspekt durch den höheren Abstraktionsgrad der SEM gewissermaßen an Relevanz. Schließlich werden bei ausreichender Stichprobengröße auch kleinste Effekte statistisch signifikant, selbst wenn sie absolut betrachtet trivial sind (Kline, 2005).

Doch nicht nur der höhere Abstraktionsgrad ist als Vorteil der SEM gegenüber anderen Verfahren zu bewerten. Ein weiterer gewichtiger Aspekt ist die Tatsache, dass im Rahmen einer Strukturgleichungsmodellierung stets auch Messfehler im Modell integriert werden und somit in der Schätzung eine Berücksichtigung finden. Es ist davon auszugehen, dass keine empirische Testung messfehlerfrei erfolgt, eine Tatsache, die beispielsweise in Regressionsanalysen unberücksichtigt bleibt. In der SEM werden Fehlerterme als latente Faktoren mit modelliert, so dass die Parameterschätzungen des Modells nicht von Fehlervarianzen beeinträchtigt werden

(Weiber & Mülhhaus, 2010). Weiterhin können Interdependenzen zwischen einzelnen Prädiktoren modelliert werden, anders als im regressionsanalytischen Ansatz, in welchem implizit die Unkorreliertheit der Prädiktoren angenommen wird (Marsh, 1990).

Die Durchführbarkeit der Strukturgleichungsmodellierung ist jedoch an gewisse Voraussetzungen gebunden, welche im folgenden Abschnitt dargestellt werden.

3.5.2.1 Voraussetzungen für die Durchführbarkeit der SEM

Die Durchführbarkeit von Strukturgleichungsanalysen ist an verschiedene Bedingungen geknüpft, die sowohl theoretischer als auch praktischer Natur sind. Wie bereits angesprochen ist aus theoretischer Sicht unabdingbar, dass im Vorfeld der Modellierung Hypothesen generiert werden, die die Zusammenhänge der untersuchten Variablen spezifizieren. Aus praktischer Sicht ist es erforderlich, die Daten und ihre Struktur vorab hinsichtlich bestimmter Merkmale zu prüfen. So verlangen zahlreiche gängige Schätzalgorithmen, wie die in dieser Arbeit eingesetzte *Maximum-Likelihood-Schätzung* (ML) multinormalverteilte Daten (Weiber & Mülhhaus, 2010). Wird diese Voraussetzung verletzt, können Verzerrungen hinsichtlich der Schätzung der Modellgüte oder der Parameter auftreten, wodurch die Gefahr falscher Schlussfolgerungen gegeben ist. Es ist jedoch zu beachten, dass der in der sozialwissenschaftlichen Forschung übliche Einsatz von Ratingskalen meist zu einer Verletzung der strengen Kriterien der Normalverteilungsannahme führt. Laut Weiber & Mülhhaus (2010) ist jedoch erst bei sehr deutlichen Verletzungen von verzerrten Schätzungen auszugehen, weshalb die ML-Schätzung auch bei moderaten Verletzungen der Verteilungsannahme genutzt werden kann (Bollen, 1989).

Üblicherweise erfolgt die Prüfung der Normalverteilungsannahme in zwei Schritten. Zunächst wird die univariate Normalverteilung der einzelnen Variablen geprüft, meist durch Inspektion der Wert für Schiefe und Kurtosis. In einem zweiten Schritt erfolgt dann die Prüfung der multivariaten Normalverteilung aller Konstrukte.

Es ist umstritten, welche Werte eine Verletzung der Normalverteilungsannahme in bedeutsamem Umfang indizieren. Nach West, Finch und Curran (1995) liegt erst ab Werten von Schiefe $> |2|$ und Kurtosis $> |7|$ eine ernste Verletzung vor.

Univariat normalverteilte Variablen stellen eine notwendige, wenngleich keine hinreichende Bedingung multinormalverteilter Daten dar (Weiber & Mülhhaus, 2010).

Daher wird üblicherweise die multivariate Normalverteilungsannahme mit Hilfe des Mardia-Testes überprüft, welcher einen multivariaten Wölbungs- Koeffizienten ermittelt. Weicht der ermittelte Koeffizient signifikant von null ab, liegen keine multivariat normalverteilten Daten vor. Häufig werden zur Überprüfung der Signifikanz *Critical Ratios* (c.r.-Werte) errechnet, durch Dividieren des ermittelten Wertes durch den jeweiligen Standardfehler. Je nachdem, ob ein strenger oder ein moderater Bewertungsmaßstab angelegt wird, sollten sich c.r.-Werte < 1.96 bzw. < 2.57 ergeben (Weiber & Mülhhaus, 2010). Nach Browne (1982; zitiert nach Weiber & Mülhhaus, 2010) führt innerhalb der Strukturgleichungsmodellierung eine stark von null abweichende Kurtosis eher zu Verzerrungen verglichen mit einer von der Normalverteilung abweichenden Schiefe, weshalb sich meist auf die Überprüfung der multivariaten Kurtosis beschränkt wird.

Weiterhin ist die Identifikation von univariaten und multivariaten Ausreißern sehr bedeutsam, da diese eine Verletzung der multivariaten Normalverteilungsannahme verursachen können. Weiber & Mülhhaus (2010) verstehen unter Ausreißern solche Werte, die „aus sachlogischer Sicht als ungewöhnlich, nicht plausibel und widersprüchlich zu bezeichnen sind und dadurch nicht zu den übrigen Beobachtungswerten einer Variablen bzw. deren Verteilung passen“ (S. 145). Dabei bezieht sich der Begriff univariat auf das Auftreten einzelner ungewöhnlicher Werte innerhalb eines Konstrukts, während multivariate Ausreißer auffällige Wertekombinationen über mehrere Variablen hinweg darstellen. Häufig wird als Cut-Off-Wert zur Identifikation univariater Ausreißer ein Wert von mehr als drei Standardabweichungen ober- bzw. unterhalb des Mittelwerts einer Variablen angegeben. Bühner (2006) weist jedoch darauf hin, dass neben der statistischen Inspektion immer auch inhaltliche Überlegungen eine Rolle spielen sollten, wenn es um die Entscheidung geht, ob Ausreißer aus dem Datensatz zu entfernen sind oder nicht. Eine rein auf statistischen Werten basierende Entscheidung ist nicht ratsam. Die Identifikation multivariater Ausreißer wird durch Betrachtung der *Mahalanobis-Distanz* ermöglicht. Hierbei wird für jeden Wert geprüft, wie weit eine Abweichung vom multivariaten Mittelwert vorliegt. Bei ausreichend großen Stichproben verteilen sich die Distanzwerte gemäß einer χ^2 -Verteilung, bei der die Anzahl der Freiheitsgrade der Anzahl der Variablen entspricht. Der Vergleich der Distanzwerte mit einem kritischen χ^2 -Wert ($p < .001$) ermöglicht die Identifikation auffälliger Fälle (Kline, 2005).

Weitere Voraussetzungen, die mit der Normalverteilungsannahme verknüpft sind, sind die Forderung nach einer linearen Beziehung zwischen den untersuchten Variablen und nach Homoskedastizität der Daten. Ob zwischen zwei Variablen eine lineare Beziehung besteht, lässt sich üblicherweise durch Inspektion der Verteilungsdiagramme prüfen, indem die jeweiligen Residuen gegeneinander geplottet werden und überprüft wird, ob sich eine zufällige Punktwolke abzeichnet. Homoskedastizität bedeutet, dass sich die Residuen zufällig entlang den Werten der abhängigen Variable verteilen. Die Fehlervarianz sollte somit für alle Werte der unabhängigen Variablen gleich bleiben (Kline, 2005). Allerdings haben auch hier leichte bzw. moderate Verletzungen der Annahme nur geringfügige Auswirkungen. Auch diese Voraussetzung lässt sich anhand von Scatterplots überprüfen. So sollte das Plotten der Werte der abhängigen Variable auf der X-Achse gegen die standardisierten vorhergesagten Werte auf der Y-Achse ein Bild ergeben, bei dem sich die Werte in gleichem Maße entlang der Regressionsgeraden verteilen. Es sei jedoch erwähnt, dass Heteroskedastizität nicht notwendigerweise durch Non-Normalität der Daten verursacht sein muss. Ebenso kann mangelnde Reliabilität eine Rolle spielen, wenn beispielsweise ein kognitiver Test sehr reliabel am oberen Verteilungsende misst, aber unreliabel ist für das untere Fähigkeitsspektrum (Kline, 2005).

Bezüglich der Beziehung der Variablen gilt es, Multikollinearität zu vermeiden. Dies bedeutet, dass die Korrelation zwischen den untersuchten Variablen nicht zu hoch ausfallen darf ($r > .85$), da ansonsten instabile Schätzungen resultieren können oder eine Modellschätzungen gar unmöglich wird (Kline, 2005).

Schließlich ist zu berücksichtigen, dass Strukturgleichungsanalysen eine gewisse Stichprobengröße voraussetzen, da zu kleine Stichproben instabile Kovarianzen bzw. Korrelationen liefern, welche als Grundlage der Modellschätzung fungieren (Kline, 2005; Tabachnick & Fidell, 2007). Nach Kline (2005) sind Stichprobengrößen kleiner 100 als klein, solche zwischen 100 und 200 als mittel und Stichproben größer 200 als groß zu bezeichnen, wobei kleine Stichproben nur für sehr einfache Modellschätzungen ausreichen. Je komplexer ein Modell, umso bedeutender wird die Größe der Stichprobe. Zwar existiert auch hier keine feste Regel, doch Kline (2005) empfiehlt ein Verhältnis von Fällen zu geschätzten Parametern von mindestens 5:1, besser 10:1. Unterhalb dieser Grenze ist die Genauigkeit der Schätzung in Frage zu stellen.

Ein weiterer wichtiger Aspekt betrifft den Umgang mit fehlenden Werten. Diese können Auswirkungen auf die Modellschätzung haben. Außerdem setzen viele Schätzalgorithmen vollständige Datensätze voraus (Weiber & Mülhhaus, 2010). Es gilt zu beachten, dass weniger die absolute Anzahl fehlender Werte eine Rolle spielt, als vielmehr die Frage, ob sich Regelmäßigkeiten im Muster fehlender Werte erkennen lassen. Daher ist zu überprüfen, ob die fehlenden Werte zufällig verteilt sind, oder ob eine Systematik vorliegt. Schließlich schränken nicht zufällig verteilte Missings die Generalisierbarkeit der Befunde ein. Im Zuge statistischer Verfahren kann nun untersucht werden, ob die fehlenden Werte vollständig zufällig (*missing completely at random*, MCAR) vorliegen, zufällig (*missing at random*, MAR), oder systematisch (*not missing at random*, NMAR). Während NMAR bedeutet, dass die fehlenden Werte eine Regelmäßigkeit erkennen lassen, bedeutet MCAR, dass es innerhalb der fehlenden Werte bzw. zwischen fehlenden und beobachteten Werten keinerlei Zusammenhänge gibt. Sind die fehlenden Werte MAR meint dies, dass es zwar einen Zusammenhang gibt zwischen der Auftretenswahrscheinlichkeit eines Missings in Variable x und den Angaben bei einer oder mehreren Variablen y, die Wahrscheinlichkeit jedoch nicht von Variable x selbst abhängt (Weiber & Mülhhaus, 2010). Erweisen sich die fehlenden Werte als MAR bzw. MCAR, ergeben sich verschiedene Möglichkeiten der Ersetzung dieser durch geschätzte Werte. MCAR liegt dann vor, wenn im Little's MCAR Test ein nicht signifikantes Ergebnis vorliegt. Erweist sich der Wert als signifikant, wird durch Inspektion der verschiedenen t-Statistiken zwischen Itempaarungen überprüft, ob die Annahme einer MAR haltbar ist oder nicht.

In der vorliegenden Arbeit wurden Missings innerhalb der Motivationsskalen mittels des *EM-Algorithmus* (Expectation Maximization) ersetzt, einem der bekanntesten Verfahren, welches hervorragende Eigenschaften aufweist (Malhotra 1987, zitiert nach Weiber & Mülhhaus, 2010). Diese iterative Prozedur ersetzt fehlende Werte in zwei Schritten. Zunächst erfolgt die Schätzung (Estimation) der fehlenden Werte durch eine Abfolge von Regressionen. In einem zweiten Schritt wird eine Maximum-Likelihood Schätzung (Maximization) des vervollständigten Datensatzes durchgeführt. Es erfolgt eine Wiederholung dieser Schritte, bis eine stabile Lösung im Maximierungsschritt resultiert (Kline, 2005). Bei der Ersetzung fehlender Werte sind stets auch inhaltliche Aspekte zu beachten. So wurden die Missings der motivationalen Skalen nur dann ersetzt, wenn der Fragebogen bearbeitet wurde, nicht jedoch, wenn kein Fragebogen abgegeben wurde und somit ein vollständiger Datensatz hätte geschätzt werden müssen.

3.5.2.2 *Modellspezifikation und –schätzung*

Wie bereits beschrieben, erfolgt die SEM in mehreren Schritten, beginnend mit der Spezifizierung eines Modells. Hierbei werden die hypothetischen Beziehungen zwischen latenten und manifesten Variablen festgelegt und in Form eines Pfadmodells graphisch dargestellt (Kline, 2005). Üblicherweise erfolgt die Darstellung latenter Variablen in Form von Ellipsen, während manifeste Variablen durch Rechtecke abgebildet werden. Fehlerterme werden durch Kreise symbolisiert. Pfeile spiegeln die kausalen Beziehungen zwischen den Variablen wider. Dabei stehen gekrümmte doppelspitzige Pfeile für ungerichtete Zusammenhänge, einfache gerade Pfeile für gerichtete Beziehungen (Tabachnik & Fidell, 2007). Eine tabellarische Übersicht der üblichen Notation ist Anhang A zu entnehmen.

Allgemein wird zwischen Messmodell und Strukturmodell differenziert. Im Messmodell wird festgelegt, welche manifesten Variablen als Indikatoren latenter Variablen herangezogen werden. Es wird also spezifiziert, „wie einer latenten Variable bzw. einem hypothetischen Konstrukt ein beobachteter Sachverhalt zugewiesen (= Operationalisierung) und durch Zahlen erfasst (= Messung) werden kann“ (Weiber & Mülhhaus, 2010, S.35). Die latente Variable soll nun die Varianz der Indikatorvariablen vorhersagen können, unter Berücksichtigung des Messfehlers (Bühner, 2006). Das Strukturmodell hingegen spezifiziert die angenommenen Beziehungen zwischen latenten Variablen, wobei *endogene* Variablen (latente abhängige Variable) von *exogenen* (latente unabhängige) Variablen unterschieden werden. Erstere werden durch andere Variablen beeinflusst und sind somit als Kriteriumsvariablen zu charakterisieren. Die in der Terminologie der SEM als exogene Variable bezeichnete Prädiktorvariable dient der Erklärung einer endogenen Variablen. In Fällen, in denen eine Variable gleichzeitig die Funktion von Kriterium und Prädiktor übernimmt, wird der Term *intervenierende Variable* angewandt (Weiber & Mülhhaus, 2010). Von großer Wichtigkeit ist, dass jeder endogenen Variable ein Messfehlerterm zuzuweisen ist, da die Varianz einer abhängigen Variablen nie vollständig durch die Prädiktoren aufgeklärt werden kann (Tabachnik & Fidell, 2007). Abbildung 5 stellt die formale Unterscheidung von Messmodell der latenten exogenen bzw. der latenten endogenen Variablen und Strukturmodell dar.

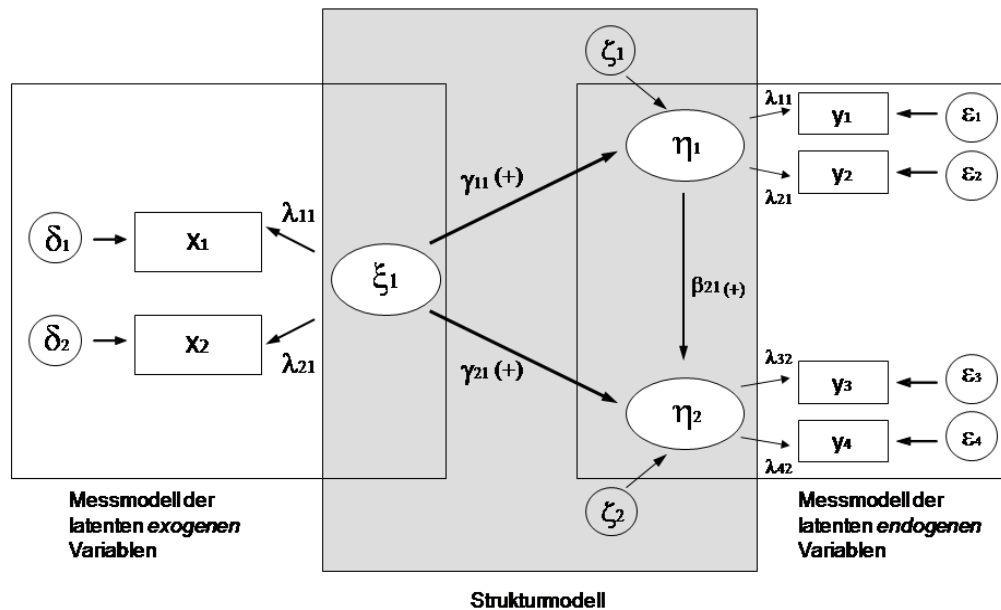


Abbildung 5: Darstellung von Mess- und Strukturmodell der Strukturgeleichungsanalyse (Weiber & Mühlhaus, 2010)

Die Anwendung dieser Terminologie auf die vorliegende Untersuchung ergibt, dass Schulerfolg als Kriterium die einzige endogene Variable darstellt, während WM, Intelligenz, SES, FSK und IM als exogene Variablen zu bezeichnen sind.

Im Anschluss an die Voraussetzungsprüfung und die Spezifikation des Modells, erfolgt die Erstellung eines linearen Gleichungssystems, wodurch das angenommene Hypothesensystem in ein lineares Mehrgleichungssystem überführt wird (Weiber & Mühlhaus, 2010). Es wird für jede abhängige Variable genau eine Gleichung formuliert. Bevor nun die eigentliche Parameterschätzung erfolgen kann, muss zunächst die Identifizierbarkeit des Modells geprüft werden. Hierbei wird geklärt, ob die erhobenen Daten genügend Informationen enthalten, um eine Parameterschätzung vornehmen zu können (Weiber & Mühlhaus, 2010). Ein Modell gilt dann als gerade identifiziert, wenn nur eine einzige Lösung der Parameterschätzung existiert (Kline, 2005). In diesem Fall entsprechen sich die Anzahl bekannter und zu schätzender Parameter ($df = 0$). Problematisch an dieser Situation ist, dass keine Modelltestung mehr möglich ist aufgrund der Tatsache, dass die Anzahl der Freiheitsgrade gleich null ist. Übersteigt die Anzahl der zu schätzenden Parameter die Zahl der bekannten

Parameter, liegt ein unteridentifiziertes Modell vor, was bedeutet, dass es sehr viele Lösungsmöglichkeiten gibt. Übersteigt nun die Anzahl bekannter Parameter die Zahl der zu schätzenden Parametern, wird von einem überidentifizierten Modell gesprochen, für welches nur eine näherungsweise Lösung gefunden werden kann, wobei die Minimierung der Diskrepanz zwischen beobachteter und implizierter Kovarianzmatrix intendiert ist (Bühner, 2006).

Ist die Frage der Identifizierbarkeit positiv beantwortet, kann schließlich die Parameterschätzung erfolgen. Diese wird im folgenden Abschnitt erläutert.

Zur Schätzung der Wirkungskoeffizienten (Pfadkoeffizienten) stehen verschiedene Algorithmen zur Verfügung, welche auf unterschiedlichen Verteilungsannahmen beruhen und verschiedene Diskrepanzfunktionen aufweisen, die es zu minimieren gilt. Am häufigsten findet jedoch die Maximum Likelihood (ML) –Schätzung Anwendung. Diese iterative Schätzmethode liefert simultane Parameterschätzungen, welche die Wahrscheinlichkeit maximieren, dass die empirische Kovarianzmatrix aus einer Population stammt, in welcher die modellimplizierte Kovarianzstruktur Validität besitzt (Schermelleh-Engel et al., 2003). Ziel dieses Vorgehens ist das Maximieren der gemeinsamen Wahrscheinlichkeitsfunktion. Die mit Hilfe der ML-Methode geschätzten Parameter und Standardfehler stellen asymptotisch unverzerzte, effiziente und konsistente Schätzungen dar, welche zur Kalkulation der modellimplizierten Kovarianzmatrix genutzt werden. Durch Bildung der Differenz zwischen den auf diesem Wege ermittelten und den empirischen Werten wird die Überprüfung der Modellgüte ermöglicht (Reinecke, 2005). Ein Vorteil der ML-Schätzung ist, dass eine statistische Prüfung des Gesamtfits eines Modells durchgeführt werden kann (siehe Abschnitt 3.5.2.3). Außerdem sind die Schätzungen skaleninvariant und skalenfrei (Bollen, 1989), weshalb es unerheblich ist, ob transformierte oder nicht transformierte Daten vorliegen oder ob Korrelations- oder Kovarianzmatrizen untersucht werden (Schermelleh-Engel et al., 2003). Als Nachteil wird häufig angemerkt, dass die ML-Schätzung positiv definite und multivariat normalverteilte Daten erfordert. Doch Simulationsstudien belegten, dass selbst bei schwerer Verletzung der Normalverteilungsannahme die Parameterschätzung konsistent blieb, so dass geschlossen werden kann, dass die ML-Methode sehr robust gegen die Verletzung der Normalverteilungsannahme ist (West, Finch & Curran, 1995; Muthén & Muthén, 2002; Curran, West & Finch, 1996). Am Ende des iterativen Schätzprozesses kann die Modellgüte geprüft werden. Hierauf wird im Folgenden eingegangen werden.

3.5.2.3 *Bewertung der Modellgüte*

Wie bereits erwähnt, liefert die Modellgüte Auskunft über das Ausmaß der Passung von empirischer und modellimplizierter Kovarianzmatrix. Die statistische Beurteilung erfolgt anhand verschiedener Fit-Indizes. Allerdings mangelt es an festen Richtlinien, welche minimalen Kriterien erfüllt sein müssen, um von einem guten Modellfit sprechen zu können, weshalb stets mehrere Fit-Indizes zur Beurteilung herangezogen werden sollten. Die Beurteilung der Gültigkeit eines Modells anhand eines einzigen statistischen Signifikanztests ist nicht möglich (Schermelleh-Engel et al., 2003). Allgemein sollte vor der Sichtung der Fit-Indizes sichergestellt werden, dass das Modell indentifiziert ist und der iterative Schätzprozess konvergiert. Außerdem sollten alle Schätzungen Werte in einem sinnvollen Bereich erreichen und die Größe der Standardfehler der Parameter keine Auffälligkeiten aufweisen (Marsh & Grayson, 2005). Es stehen nun eine Vielzahl deskriptiver Indizes zur Verfügung, wobei die Entscheidung, anhand welcher Werte die Gütebeurteilung erfolgt, keineswegs trivial ist (Schermelleh-Engle et al., 2003).

Nach Schermelleh-Engel et al. (2003) ist der χ^2 - Test der einzige inferenzstatistische Fit-Index, welcher die Gültigkeit des Gesamtmodells anhand eines Signifikanztests bewertet. Dabei wird die Nullhypothese geprüft, dass die beobachtete und modellimplizierte Kovarianzmatrix kompatibel sind. Je größer die Unterschiede, umso höhere χ^2 -Werte ergeben sich. Ein mit dem jeweiligen χ^2 -Wert assoziierter p-Wert von $p > .05$ hingegen bestätigt die Nullhypothese und führt zu dem Schluss, dass das Modell auf die empirischen Daten passt (Schermelleh-Engel et al., 2003). Problematisch am χ^2 -Test ist jedoch die Stichprobenabhängigkeit. So steigt der χ^2 – Wert bei steigender Stichprobengröße (und konstanten Freiheitsgraden) an, was zur Folge haben kann, dass ein Modell selbst bei trivialer Diskrepanz abgelehnt wird (Kline, 2005). Umgekehrt ist es möglich, dass bei kleinen Stichproben Modelle trotz einer substantiellen Abweichung zwischen beiden Kovarianzmatrizen akzeptiert werden, da aufgrund der sinkenden Stichprobengröße auch der χ^2 -Wert sinkt und somit leichter ein nicht signifikanter Wahrscheinlichkeitslevel erreicht wird (Schermelleh-Engel et al., 2003). Der χ^2 -Wert ist außerdem abhängig von der Komplexität eines Modells. So sinkt der Wert mit steigender Parameterzahl aufgrund der sinkenden Anzahl an Freiheitsgraden (Schermelleh-Engel et al., 2003). Aus diesen Gründen wird empfohlen, die χ^2 -Statistik lediglich als deskriptives Gütekriterium anzusehen und nicht im Sinne einer Inferenzstatistik. Jöreskog &

Sörbom (1993) raten weiterhin an, das Verhältnis von χ^2 -Wert und Anzahl der Freiheitsgrade (χ^2/df) zu inspizieren. Werte nahe zwei indizieren einen guten und Werte um drei einen akzeptablen Fit (Schermelleh-Engel et al., 2003). Zusätzlich sollten stets alternative Bewertungskriterien der Modellgüte betrachtet werden. Die in der vorliegenden Arbeit berücksichtigten Kennwerte der Anpassungsgüte (*Goodness of fit*) werden im folgenden Abschnitt thematisiert. Es sind dies GFI, CFI, RMSEA und AIC.

Ein Großteil der Fit-Indizes wurde aus dem χ^2 -Test abgeleitet und ist deskriptiver Natur. Diese Kennwerte evaluieren entweder die Gesamtgüte des Modells, die Güte eines Modells im Vergleich zu Alternativmodellen oder die Modellsparsamkeit (Schermelleh-Engel et al., 2003).

Der Goodness-of-Fit-Index (GFI; Jöreskog & Sörbom, 1989) testet das zu schätzende Modell gegen das restriktive Nullmodell, in welchem sämtliche Parameter auf null fixiert vorliegen. Der χ^2 -Wert des Testmodells und der χ^2 -Wert des Nullmodells werden hierzu ins Verhältnis gesetzt (Schermelleh-Engel et al., 2003). Dieser Kennwert kann Werte zwischen null und eins annehmen, wobei höhere Werte einen besseren Fit anzeigen. Während Werte von $\text{GFI} > .90$ einen akzeptablen Fit indizieren, gelten Werte von $\text{GFI} > .95$ als indikativ für einen guten Fit.

$$\text{GFI} = 1 - \frac{\chi_t^2}{\chi_n^2}$$

χ_t^2 = Chi-Quadrat des Nullmodells

χ_n^2 = Chi-Quadrat des untersuchten Modells (Targetmodell)

Der *Comparative Fit-Index* (CFI; Bentler, 1990) bewertet die Güte des Modells im Vergleich zum *Independence model*. Dieses Baselinemodell weist auf null fixierte Parameter auf, mit Ausnahme der Varianzen der manifesten Variablen. Der CFI indiziert, ob das zu testende Modell eine Verbesserung gegenüber dem Independence model darstellt (Schermelleh-Engel et al., 2003). Der CFI variiert zwischen null und eins, wobei ein besserer Fit durch höhere Werte angezeigt wird. Als „Daumenregel“ gilt, dass Werte von $\text{CFI} > .97$ einen guten Modellfit und Werte von $\text{CFI} > .95$ einen akzeptablen Fit bedeuten (Schermelleh-Engel et al., 2003). Die Formel zeigt, dass dieser Index die χ^2 - Werte des Baseline- und des Testmodells sowie die Anzahl der Freiheitsgrade berücksichtigt.

$$CFI = 1 - \frac{\max[(\chi_t^2 - df_t), 0]}{\max[(\chi_t^2 - df_t), (\chi_i^2 - df_i), 0]}$$

χ_i^2 = Chi-Quadrat des Independence Model

χ_t^2 = Chi-Quadrat des untersuchten Modells (Targetmodell)

df = Anzahl der Freiheitsgrade

Der *Root Mean Square Error of Approximation* (RMSEA; Steiger, 1990) ist ein deskriptiver Index, der die Güte des Gesamtmodells beschreibt. Im Gegensatz zum χ^2 -Test ist der RMSEA weniger von der Stichprobengröße abhängig. Außerdem prüft er nicht den exakten Fit beider Kovarianzmatrizen, sondern die realistischere Annahme einer guten Annäherung, was als „close fit“ (Browne & Cudeck, 1993, S. 146) bezeichnet wird. Desweiteren berücksichtigt der RMSEA die Komplexität eines Modells und belohnt parsimonischere Modellierungen (Weiber & Mülhhaus, 2010). Allgemein weisen kleinere Werte auf einen besseren Fit hin. Ein sehr guter Fit wird durch Werte von RMSEA < .01 und ein guter Fit durch Werte < .05 angezeigt. Werte von RMSEA < .08 indizieren einen akzeptablen Fit (Neale, Lubke, Aggen & Dolan, 2005). Außerdem sollte das 90% -ige Konfidenzintervall den Wert null einschließen (Schermele-Engel et al., 2003). Wie die Formel zeigt, ergibt sich der RMSEA aus dem χ^2 -Wert, der Anzahl der Freiheitsgrade (df), sowie der Anzahl der Probanden (n).

$$RMSEA = \sqrt{\frac{\chi^2 - df}{df (n - 1)}}$$

χ^2 = Chi-Quadrat des untersuchten Modells

df = Anzahl der Freiheitsgrade

n = Anzahl der Probanden

Eine tabellarische Darstellung gängiger „Daumenregeln“ zur Klassifizierung der Modellgüte kann Anhang A entnommen werden.

Weisen die Gütekriterien auf eine schlechte Passung des Modells hin, können Änderungen vorgenommen werden, die sich durch Sichtung der sogenannten Modifikationsindizes ergeben. Diese helfen einzelne Fehlspezifikationen zu identifizieren, doch auch hier gilt, dass solche post-hoc Spezifikationen nur dann

zulässig sind, wenn eine theoretische oder inhaltliche Sinnhaftigkeit gegeben ist (Bühner, 2006).

Bewertung von Alternativmodellen

Die Beurteilung der Modellgüte ist jedoch nicht nur bei der Bewertung eines einzelnen Modells relevant. Besonders wenn es darum geht zwischen Modellalternativen zu entscheiden, muss auf Kriterien der Modellgüte verwiesen werden. Dabei erfolgt eine Differenzierung zwischen *genesteten* Modellen, welche hierarchisch strukturiert sind und verschiedene Restriktionen aufweisen und *nicht genesteten* Modellen, welche unabhängig voneinander geschätzt werden. Zwei Modelle sind dann genestet, wenn sie sich durch Freisetzen, Fixieren oder Einfügen einer anderen Form von Restriktion mindestens eines Parameters ineinander überführen lassen (Schermelleh-Engel et al., 2003). Der Vergleich genesteter Modelle erfolgt anhand des χ^2 -Differenztests, welcher die Differenz im Modellfit testet. Die Teststatistik jedes einzelnen Modells folgt der χ^2 -Verteilung, so dass sich auch χ^2 -verteilte Differenzwerte ergeben (Steiger, Shapiro & Browne, 1995). Die Anzahl der Freiheitsgrade ergibt sich aus der Differenz der Freiheitsgrade der beiden Modelle, die es zu vergleichen gilt. Die gemäß der abgebildeten Formel errechnete χ^2 - Differenz kann auf Signifikanz geprüft werden. Fittet das genestete Modell nicht signifikant schlechter ($p > .05$) als das vollständige Modell, bestätigt dies die Nullhypothese, dass beide Modelle eine gleich gute Passung aufweisen. In diesem Fall wird das sparsamere, genestete Modell dem Parsimonitätskriterium folgend dem vollständigen Modell vorgezogen (Neale & Cardon, 1992). Allerdings ist zu beachten, dass die Stichprobenabhängigkeit der χ^2 - Wertes auch bei genesteten Modellen bestehen bleibt (Schermelleh-Engel et al., 2003).

$$\chi^2_{\text{diff}} (\text{df}_{\text{diff}}) = \chi^2_{\text{A}} (\text{df}_{\text{A}}) - \chi^2_{\text{B}} (\text{df}_{\text{B}})$$

$$\text{df}_{\text{diff}} = \text{df}_{\text{A}} - \text{df}_{\text{B}}$$

A: restringierte Version des Modells B (genestetes Modell)

B: vollständiges Modell

(Bentler, 1990; Bollen, 1989; Jöreskog, 1993)

Sollen zwei nicht genestete Alternativmodelle verglichen werden, muss auf andere Kennwerte zurückgegriffen werden. Ein weit verbreiteter deskriptiver Index, der das

Parsimonitätsprinzip berücksichtigt, ist das *Akaike Information Criterion* (AIC; Akaike, 1974, 1987). Dieser Kennwert berücksichtigt χ^2 -Wert und Anzahl der Freiheitsgrade des Modells. Das Modell, welches den kleineren AIC-Wert aufzuweisen hat, gilt als passender. Es ist jedoch zu beachten, dass es sich beim AIC um einen nicht normierten Index handelt, so dass der Wert stets im Vergleich zu einem anderen AIC-Wert betrachtet werden muss und nicht für sich selbst interpretiert werden kann (Schermelleh-Engel et al., 2003).

$$\text{AIC} = \chi^2 - 2df$$

$$\chi^2 = \chi^2 - \text{Wert}$$

df = Anzahl der Freiheitsgrade

Abschließend bleibt anzumerken, dass trotz der beachtlichen Stärken der SEM stets auf eine angemessene Interpretation der Ergebnisse zu achten ist. Ein guter Modellfit ist nicht gleichzusetzen mit der Schlussfolgerung, der Nicht-Existenz von Alternativmodellen, die eine gleich gute, wenn nicht bessere Passung aufweisen (Confirmation Bias). Die Gültigkeit eines Modells bezieht sich außerdem stets auf die erhobene Stichprobe, so dass eine gewisse Vorsicht anzuraten ist bezüglich der Generalisierung der Befunde auf eine Population.

IV Ergebnisse

Im folgenden Kapitel werden zunächst die deskriptiven Statistiken aller Konstrukte dargestellt, sowie verschiedene korrelative Zusammenhänge erläutert (Kapitel 4.1). Anschließend erfolgt die Darstellung der Regressions- und Kommunalitätenanalysen der kognitiven Prädiktoren, getrennt für Deutsch und Mathematik (Kapitel 4.2.1). Im Anschluss daran werden die Befunde der Strukturgleichungsmodellierung dargestellt. Hierbei werden zunächst die Modelle berichtet, in welche lediglich WM, gf und SES als Prädiktoren eingingen (Kapitel 4.2.2). Danach werden die Ergebnisse der Modellierungen dargestellt, die gleichzeitig kognitive und non-kognitive Faktoren, sowie den SES berücksichtigen (Kapitel 4.3).

4.1 Deskriptive Statistiken und korrelative Zusammenhänge

4.1.1 Deskriptive Statistiken und Binnenstruktur der kognitiven Prädiktoren

Die Verteilungskennwerte und Reliabilitäten der drei Arbeitsgedächtnisaufgaben und CFT-Subtests sind Tabelle 4 zu entnehmen. Alle Arbeitsgedächtnisaufgaben wiesen eine geringfügige Linksschiefe auf (-.03 bis -.46), ebenso die Subtests des CFT (-.16 bis -.55). Lediglich CFT-Subtest 1 wies mit einer Schiefe von .03 praktisch keine Abweichung von der Schiefe einer Normalverteilung auf. Bezüglich der Kurtosis ergab sich, dass DS (.27), CFT-Subtests 3 (.16) und CFT-Subtest 4 (.11) leicht steilgipflige Verteilungen aufwiesen, während sich MS (-.53), AS (-.26), CFT-Subtest 1 (-.47) und Subtest 2 (-.30) als eher flachgipflig im Vergleich zur Normalverteilung darstellten. Da jedoch weder die Arbeitsgedächtnisaufgaben noch die CFT-Subtests Werte von über $|1|$ für Schiefe und Kurtosis aufwiesen, konnte trotz signifikanter Werte im Kolmogorov-Smirnov-Test (korrigiert nach Lilliefors) von annähernd normalverteilten Skalen ausgegangen werden. Auch die Inspektion der graphischen

Verteilungsdiagramme⁸ der vier CFT-Subtests bzw. der drei WM-Tasks, ergab keine Hinweise auf auffällige Verteilungen.

Da es sich bei den Arbeitsgedächtnisaufgaben nicht um standardisierte Verfahren handelte, sondern um Adaptationen bestehender Tests, wurde der Range der kognitiven Tests einer genaueren Betrachtung unterzogen, um Boden- bzw. Deckeneffekte auszuschließen. Es ergab sich, dass 9.7% der Gesamtstichprobe (N = 31) alle zehn Trials der DS-Task richtig gelöst hatten. Zwei Personen (0.6%) lösten keinen Trial richtig. Kein Proband konnte alle zehn AS-Trials richtig lösen. Drei Personen (0.9%) bearbeiteten die MS-Task fehlerfrei. Über alle drei Aufgaben hinweg lag das Maximum an richtig gelösten Aufgaben, welches von einer Person erzielt wurde, bei 28 von 30 Aufgaben. Das Minimum lag bei insgesamt sieben richtig gelösten Aufgaben, welches ebenfalls bei einer Person vorlag. Insgesamt ergab sich also für keine Arbeitsgedächtnisaufgabe der Verdacht eines Boden- oder Deckeneffektes. Zwar zeigte sich in der DS-Task der größte Anteil an Personen mit voller Punktzahl, doch auch hier wurde unter Berücksichtigung der Werte für Schiefe und Kurtosis eine Ergebnisverzerrung ausgeschlossen.

Hinsichtlich der CFT- Subskalen zeigte sich, dass ein Proband Subtest 1 vollständig richtig lösen konnte. Zwei weitere Personen (0,6%) bearbeiteten alle 15 Aufgaben des Subtests 3 fehlerfrei. Die Subtests 2 und 4 wurden von keiner Person ohne Fehler gelöst. Sechs Probanden (1.9%) lösten keine Aufgabe des vierten Subtests richtig. Insgesamt sprachen die Befunde auch hier gegen das Vorliegen von Boden- bzw. Deckeneffekten im CFT 20-R.

Die Betrachtung univariater Extremwerte (siehe Kapitel 3.5.2) ergab bezüglich der Arbeitsgedächtnisaufgaben keinerlei Auffälligkeiten mit Ausnahme von zwei Ausreißern in der DS-Task. Im CFT fanden sich zwei Extremwerte in Subtest 1. Einer dieser Fälle trat auch als Ausreißer in Subtest 3 in Erscheinung. Darüber hinaus fand sich in Subtest 3 ein zweiter Extremwert. Insgesamt ergaben sich somit fünf Ausreißer innerhalb der kognitiven Skalen. Die Tatsache, dass jedoch für keine Skala die 5% getrimmten Mittel von den arithmetischen Mitteln abwichen, sprach gegen eine Exklusion der Fälle. Allerdings erwiesen sich in einem Fall die schlechten Leistungen des Probanden im CFT als nicht konsistent im Hinblick auf die Leistungen

⁸ Unter graphischen Verteilungsdiagrammen werden Q-Q-Plots und trendbereinigte Plots verstanden.

in den Arbeitsgedächtnisaufgaben, da diese starke Leistungsschwankungen aufwiesen (neun richtige Lösungen in der DS-Task bei null richtigen Lösungen in der MS-Task). Dieser Fall wurde in den folgenden Analysen exkludiert, da nicht auszuschließen war, dass in der DS-Aufgabe gegen die Instruktion gearbeitet worden war. Außerdem wurde ein weiterer Fall ausgeschlossen, da auch hier ein auffälliges Antwortmuster vorlag, da bei schlechter Leistung in allen Arbeitsgedächtnisaufgaben, sowie in drei der vier CFT Subtests, neun richtige Lösungen im zweiten Subtest des CFT erzielt wurden. In den verbleibenden Extremfällen deckte sich die unterdurchschnittliche Leistung im CFT mit den Leistungen in den Arbeitsgedächtnisaufgaben, weshalb auf eine Exklusion verzichtet wurde, um ein breiteres Fähigkeitsspektrum innerhalb der Stichprobe aufrechtzuerhalten. Damit umfasste die Stichprobe in den folgenden Analysen ein N von 316 Personen.

Vor dem Hintergrund, dass es sich bei der untersuchten Stichprobe um eine sehr homogene Gruppe bezüglich des Alters handelte, war nicht von altersspezifischen Effekten auszugehen.

Als Maß für die Reliabilität der Skalen wurde die Splithalf Reliabilität berichtet, errechnet nach dem Odd-Even Prinzip (r_{tt}) (siehe Tabelle 4). Da sich dieser Wert auf die halbierte Itemzahl bezieht, wurde die Reliabilität nach Spearman-Brown korrigiert (Lienert, 1989). Der korrigierte Wert (r_{ttc}) diente als Maß für die interne Konsistenz der Skalen und variierte für die Arbeitsgedächtnismaße zwischen $r_{ttc} = .50$ und $r_{ttc} = .73$. Während die Werte der beiden Skalen DS ($r_{ttc} = .60$) und MS ($r_{ttc} = .73$) im akzeptablen Bereich lagen, erwies sich die AS-Task als unreliabler ($r_{ttc} = .50$). Über alle drei Tasks hinweg wurde ein Reliabilitätswert von $r_{ttc} = .77$ erreicht.

Die internen Konsistenzen der CFT-Subtests lagen mit Werten zwischen $r_{ttc} = .47$ bis $.67$ ebenfalls nur in einem niedrigen bis akzeptablen Bereich. Besonders Subtest 1 erwies sich mit einem Wert von $r_{ttc} = .47$ als weniger reliabel als die restlichen Subtests. Da im Manual des CFT 20-R jedoch keine Einzelreliabilitäten für Subskalen berichtet werden, wurde die Reliabilität der Gesamtskala berechnet. Es ergab sich ein Wert von $r_{ttc} = .78$, welcher sich dem im Manual berichteten Wert von $r_{tt} = .92$ für den Gesamttest annäherte (siehe Tabelle 4).

Tabelle 4: Deskriptive Statistiken der kognitiven Fähigkeiten

	Items	r_{tt}	r_{ttc}	M	SD	Min	Max	S	K	K-S
Arbeitsgedächtnis										
WMsum	30	.63	.77	17.77	3.72	7	28	.13	.13	
DS	10	.44	.60	7.45	1.58	2	10	-.46	.27	.14**
AS	10	.33	.50	5.34	1.32	2	8	-.03	-.26	.15**
MS	10	.57	.73	4.96	2.25	0	10	-.09	-.53	.10**
Intelligenz										
CFTsum	56	.64	.78	32.89	6.04	18	50	.12	-.43	
CFT 1	15	.30	.47	9.83	2.03	1	15	.03	-.47	.10**
CFT 2	15	.35	.52	7.69	2.18	1	13	-.16	-.30	.11**
CFT 3	15	.51	.67	9.95	2.48	1	15	-.55	.16	.12**
CFT 4	11	.38	.55	5.42	2.00	0	10	-.29	.11	.16**

Anmerkungen: N = 316; WMsum: Summenscore über drei WM-Tasks; DS: Digit Sorting Task; AS: Animal Sorting Task; MS: Backward Matrix Span Task; CFTsum: Summenscore über vier CFT Subtests; CFT 1 bis CFT 4: CFT Subtests; r_{tt} : Splithalf Reliabilität (odd-even); r_{ttc} : korrigierte Splithalf Reliabilität nach Spearman-Brown; M: Mittelwert; SD: Standardabweichung; Min: Minimum; Max: Maximum; S: Schiefe; K: Kurtosis; K-S: Kolmogorov-Smirnoff-Test (korrigiert nach Lilliefors).

Tabelle 5 zeigt die Binnenstruktur des Arbeitsgedächtnisses in Form von bivariaten Korrelationen der Rohpunktwerte der Arbeitsgedächtnisaufgaben, sowie ihre Ladungen auf einem gemeinsamen Arbeitsgedächtnisfaktor.

Die Korrelationen der einzelnen Arbeitsgedächtnisaufgaben mit einem gemeinsamen Arbeitsgedächtnisfaktor variierten zwischen $r = .77$ für die DS- und AS-Task und $r = .61$ für die MS-Task und waren somit insgesamt als hoch zu bezeichnen. Dies bestätigte die Annahme eines zugrunde liegenden gemeinsamen Arbeitsgedächtnisfaktors. Dieser unrotierte Faktor⁹ mit einem Eigenwert von 1.58 erklärte 52.4% der Varianz innerhalb der Arbeitsgedächtnisaufgaben. Die Werte für Schiefe und Kurtosis deuteten nicht auf eine Verletzung der Normalverteilungsannahme hin (Schiefe = .07; Kurtosis = .17).

⁹ Der WM-Faktor wurde mittels einer Hauptkomponentenanalyse berechnet.

Tabelle 5: Korrelative Betrachtung der Binnenstruktur der kognitiven Prädiktoren sowie der Interkorrelationen

	Arbeitsgedächtnis			Intelligenz					Faktorladung	
	AS	MS	WM	CFT1	CFT2	CFT3	CFT4	gf	gf	WM
DS	.39	.24	.77	.30	.20	.28	.24	.37		.78
AS	--	.22	.77	.23	.20	.22	.16	.29		.76
MS		--	.61	.23	.15	.27	.12	.29		.62
WM			--	.36	.25	.36	.25	.45		
CFT1				--	.24	.40	.32	.71	.63	
CFT2					--	.33	.29	.65	.47	
CFT3						--	.30	.75	.68	
CFT4							--	.68	.53	

Anmerkungen: N = 316; WM: unrotierter Arbeitsgedächtnisfaktor; gf: unrotierter Intelligenzfaktor; alle r signifikant ($p < .01$), außer *kursiv* = $p < .05$.

Tabelle 5 bildet außerdem die Binnenstruktur des erhobenen Intelligenzkonstrukts ab in Form von bivariaten Korrelationen der Rohwerte der CFT-Subtestskalen, sowie ihre Ladungen auf einen gemeinsamen gf -Faktor. Alle vier Subskalen korrelierten positiv und hoch signifikant, was auf einen deutlichen Anteil geteilter Varianz zwischen den Subtests hinwies und die Annahme eines gemeinsamen gf-Faktors stützte. Die niedrigste Korrelation fand sich zwischen Subtest 1 und Subtest 2 ($r = .24$), die höchste zwischen Subtest 1 und 3 ($r = .40$).

Alle vier Subtests korrelierten hoch signifikant positiv mit einem unrotierten Intelligenzfaktor in einem Range von $r = .65$ bis $r = .75$. Dieser extrahierte Faktor mit einem Eigenwert von 1.94 erklärte 48.57% der Varianz und lieferte keine Hinweise auf eine Verletzung der Normalverteilungsannahme (Schiefe = .13; Kurtosis = -.41).

Neben der Binnenstruktur von WM und gf wurden außerdem die korrelativen Zusammenhänge beider Konstrukte analysiert. Auch diese sind Tabelle 5 zu entnehmen. Alle Korrelationen erwiesen sich als positiv und hoch signifikant, mit Ausnahme der Korrelation zwischen MS und CFT4 ($r = .12$), welche nur auf dem 5% - Niveau Signifikanz erreichte. Besonders hohe Korrelationen fanden sich zwischen DS und den CFT-Subtests ($r = .20 - .30$). Die AS-Task korrelierte in einem Range von $r = .16$ bis $r = .23$ mit den CFT-Subtests und die MS-Task zwischen $r = .12$ und $r = .27$. Alle drei Arbeitsgedächtnisaufgaben korrelierten mit $r > .29$ mit dem gf- Faktor. Die Zusammenschau der in Tabelle 4 und Tabelle 5 dargestellten Ergebnisse zeigt,

dass wie im Sinne der konvergenten Validität zu erwarten, der gf- Faktor stärker als der WM-Faktor mit den CFT Subskalen korrelierte und umgekehrt, der WM-Faktor stärker mit den WM Tasks assoziiert war als der gf-Faktor, wobei sich alle Korrelationen als positiv und hoch signifikant erwiesen. Zwischen den Faktoren gf und WM zeigte sich ein substantieller Zusammenhang von $r = .45$.

4.1.2 Deskriptive Statistiken und Binnenstruktur der non-kognitiven Prädiktoren

Die deskriptiven Statistiken der fachspezifischen FSK- und IM-Skalen sind Tabelle 6 zu entnehmen. Alle Skalen wiesen moderate bis hohe interne Konsistenzen auf ($\alpha = .65 - .81$), wobei die Skalen für Mathematik etwas reliabler ausfielen als die Deutschskalen.

Standardabweichungen und Mittelwerte fielen für alle vier Skalen in vergleichbarer Höhe aus, wobei im Fach Mathematik etwas höhere Mittelwerte erreicht wurden als im Fach Deutsch. Insgesamt fielen die Mittelwerte wie für die Altersgruppe zu erwarten war, relativ hoch aus. Sie lagen zwischen $\bar{x} = 3.79$ und $\bar{x} = 4.14$ für FSK und zwischen $\bar{x} = 3.77$ und $\bar{x} = 4.04$ für IM.

Bezüglich Schiefe und Kurtosis ist zu bemerken, dass sowohl FSK als auch IM in Mathematik mit einer Schiefe = -1 als linksschief zu bezeichnen sind. Auch die Linksschiefe der Skala IM Deutsch ist nennenswert (-.72). Die Verteilung des FSK Deutsch stellte sich als weniger linksschief dar (-.49). Mit Ausnahme der bezüglich der Kurtosis kaum von der Normalverteilung abweichenden Skala FSK Deutsch (-.24), ergaben sich für alle restlichen Skalen tendenziell steilgipflige Verteilungsmuster. So erreichte die IM-Skala im Fach Mathematik einen Wert von Kurtosis = .84. Dennoch lag die Kurtosis für alle vier Skalen unterhalb des Wertes von $|1|$, so dass dieser Kennwert keinen Hinweis auf problematische Verteilungen lieferte.

Alle vier Skalen wiesen signifikante Werte im Kolmogorov-Smirnov-Test auf, doch dies ist als Folge der Stichprobengröße zu sehen, auf welche dieser Test sehr sensibel reagiert. Außerdem weisen Tabachnik & Fidell (2001) darauf hin, dass in ausreichend großen Stichproben die Schiefe der Verteilung keinen substantiellen

Einfluss ausübt bzw. der tatsächliche Wert von größerer Bedeutung ist als die statistische Signifikanz. Auch die Inspektion der Verteilungsdiagramme aller vier Skalen lieferte keine Hinweise auf eine nennenswerte Verletzung der Normalverteilungsannahme, so dass von weiteren Prüfungen bezüglich der Normalverteilungsannahme abgesehen wurde.

Hinsichtlich des Auftretens univariater Extremwerte fanden sich jeweils zwei Fälle innerhalb der Skala IM in Deutsch und in Mathematik. Da jedoch in allen Skalen die 5% getrimmten Mittel nicht von den arithmetischen Mittelwerten abwichen, die Korrelationen zwischen den Skalen, zu den kognitiven Prädiktoren und zum Kriterium sich nicht durch Exklusion dieser Fälle veränderten und kein auffälliges Muster in den betreffenden Antworten gefunden werden konnte, wurde davon abgesehen diese vier Fälle auszuschließen.

Tabelle 6: Deskriptive Statistiken der non-kognitiven Faktoren

	Items	N	Alpha	M	SD	Min	Max	S	K	K-S
Fähigkeitsselbstkonzept										
FSKd	3	288	.65	3.79	.73	1	5	-.49	-.24	.15**
FSKm	3	285	.81	4.14	.80	1	5	-1.0	.66	.16**
Intrinsische Motivation										
IMd	3	290	.65	3.77	.81	1	5	-.72	.54	.13**
IMm	3	291	.72	4.04	.89	1	5	-1.0	.84	.16**

Anmerkungen: FSKd: Skalenwert Fähigkeitsselbstkonzept Deutsch; FSKm: Skalenwert Fähigkeitsselbstkonzept Mathematik; IMd: Skalenwert intrinsische Motivation Deutsch; IMm: Skalenwert intrinsische Motivation Mathematik; Items: Anzahl der Items pro Skala; Alpha: Cronbachs α ; M: Mittelwert; SD: Standardabweichung; Min: Minimum; Max: Maximum; S: Schiefe; K: Kurtosis; K-S: Kolmogorov-Smirnoff-Test korrigiert nach Lilliefors.

Tabelle 7 stellt sowohl die Korrelationen innerhalb der Skalen FSK und IM dar, als auch die Korrelationen zwischen den Skalen, innerhalb der gleichen Domäne und über die verschiedenen Domänen hinweg. Die konvergenten Korrelationen innerhalb der Skalen erwiesen sich als moderat. So fand sich zwischen FSK in Deutsch und Mathematik eine Korrelation von $r = .30$. Die Korrelation zwischen IM in Deutsch und Mathematik betrug $r = .22$, wobei diese Korrelation lediglich auf einem Level von $p = .05$ Signifikanz erreichte.

Die Betrachtung der Zusammenhänge zwischen den beiden Motivationsskalen ergab das Muster, dass Einschätzungen innerhalb derselben Domäne stärker zusammenhingen, als die Einschätzungen über die Domänen hinweg, da für diese keine signifikanten Korrelationen nachgewiesen werden konnten. So korrelierte FSK in Deutsch mit IM in diesem Fach ($r = .50$), nicht jedoch mit IM in Mathematik ($r = -.07$). Umgekehrt fand sich ein Zusammenhang von $r = .66$ zwischen FSK in Mathematik und der entsprechenden Einschätzung der intrinsischen Motivation, bei einer nicht signifikanten Korrelation zwischen FSK in Mathematik und IM in Deutsch ($r = .10$).

Tabelle 7: Konvergente und divergente Korrelationen der non-kognitiven Faktoren innerhalb einer Domäne und über die Domänen hinweg

	FSKd	FSKm	IMd	IMm
<i>Fähigkeitsselbstkonzept</i>				
Deutsch (FSKd)	-	.30**	.50**	.07
Mathematik (FSKm)		-	.10	.66**
<i>Intrinsische Motivation</i>				
Deutsch (IMd)				.22*
Mathematik (IMm)				-

Anmerkungen: FSKd: Skalenwert Fähigkeitsselbstkonzept Deutsch; FSKm: Skalenwert Fähigkeitsselbstkonzept Mathematik; IMd: Skalenwert intrinsische Motivation Deutsch; IMm: Skalenwert intrinsische Motivation Mathematik; * = $p < .05$; ** = $p < .01$, *kursiv* = nicht signifikant.

4.1.3 Deskriptive Statistiken und Binnenstruktur des SES

Die Beschreibung der Kategorien des höchsten elterlichen Schulabschlusses sowie deren Häufigkeitsverteilungen wurde bereits in Kapitel 3.3 vorgenommen. Ergänzend sollen nun die Verteilungskennwerte und korrelativen Zusammenhänge zwischen dem höchsten Abschluss der Väter und Mütter dargestellt werden (Tabelle 8). Während sich die Darstellung in Kapitel 3.3 aus Gründen der Vergleichbarkeit mit dem Datensatz des Datenreports Deutschland 2011 auf die verschiedenen Schulabschlüsse bezog (kein Abschluss, Hauptschulabschluss, Realschulabschluss und Fachabitur/ Abitur), wurde in den folgenden Analysen zusätzlich die Kategorie *abgeschlossenes Studium* unterschieden, um eine genauere Differenzierung des

höchsten Bildungsabschlusses zu erhalten. In folgenden Analysen wird der höchste Bildungsabschluss als Indikator des SES betrachtet.

Tabelle 8: Verteilungskennwerte und korrelative Zusammenhänge des höchsten elterlichen Bildungsabschlusses

	Min	Max	M	SD	S	K	SES _V	SES _E
Höchster Bildungsabschluss Mutter	0	5	2.78	1.39	-.40	-.14	.63	.88
Höchster Bildungsabschluss Vater	0	5	2.60	1.67	-.09	-.94	--	.83
Höchster Bildungsabschluss Eltern	0	5	3.13	1.50	-.53	-.33		--

Anmerkungen: N = 316¹⁰; Min: Minimum; Max: Maximum; M: Mittelwert; SD: Standardabweichung; S: Schiefe; K: Kurtosis; SES_V: Höchster Bildungsabschluss des Vaters; SES_E: Höchster Bildungsabschluss innerhalb eines Elternpaares; alle r signifikant ($p < .01$), außer *kursiv* = $p < .05$.

Keine Skala wies Hinweise auf problematische Verteilungen auf, da alle Werte für Schiefe und Kurtosis als unkritisch zu bezeichnen waren. Mittelwerte und Standardabweichungen wiesen für Mütter und Väter vergleichbare Werte auf, bei leicht höherem Mittelwert der Mütter und höherer Standardabweichung der Väter. Zwischen dem höchsten Bildungsabschluss der Väter und Mütter fand sich eine hoch signifikante Korrelation von $r = .63$. Sowohl der höchste Abschluss der Mütter als auch der Abschluss der Väter korrelierte erwartungsgemäß hoch mit dem höchsten Abschluss innerhalb eines Elternpaares ($r = .88 / .83$).

¹⁰ In die Kategorie *ohne Angaben* gingen sowohl die Personen ein, welche keine Auskünfte bezüglich ihres Abschlusses gaben, als auch Personen, die keinen Fragebogen bearbeitet hatten.

4.1.4 Korrelative Betrachtung der Zusammenhänge zwischen den Prädiktoren

Da in den sich anschließenden Analysen kognitive und non-kognitive Variablen, sowie der elterliche SES als konkurrierende Prädiktoren eingingen, erfolgte vorab die Betrachtung der Zusammenhänge zwischen den verschiedenen Prädiktoren. Diesbezügliche Werte sind Tabelle 9 zu entnehmen. Eine vollständige Korrelationsmatrix, die alle korrelativen Zusammenhänge der untersuchten Variablen auch auf Subtestniveau darstellt, ist in Anhang A zu finden.

Tabelle 9: Korrelative Zusammenhänge der kognitiven und non-kognitiven Prädiktoren und des SES

	FSKd	FSKm	IMd	IMm	SES _M	SES _V	SES _E
Kognitive Prädiktoren							
Arbeitsgedächtnis (WM)	.31**	.32**	.03	.16**	.14*	.12	.15**
Intelligenz (gf)	.14*	.18**	-.08	.11	.29**	.21**	.26**
Non-kognitive Prädiktoren							
FSKd					.14*	.11	.18**
FSKm					.16*	.14*	.18**
IMd					-.05	-.12*	-.04
IMm					-.01	-.03	-.01

Anmerkungen: N > 282; WM: Arbeitsgedächtnisfaktor; gf: Intelligenzfaktor; SES_M: Höchster Bildungsabschluss der Mutter; SES_V: Höchster Bildungsabschluss des Vaters; SES_E: Höchster Bildungsabschluss innerhalb eines Elternpaares; FSKd: Skalenwert Fähigkeitsselbstkonzept Deutsch; FSKm: Skalenwert Fähigkeitsselbstkonzept Mathematik; IMd: Skalenwert intrinsische Motivation Deutsch; IMm: Skalenwert intrinsische Motivation Mathematik; * = p < .05; ** = p < .01.

Wie aus Tabelle 9 ersichtlich, korrelierten nicht alle non-kognitiven Prädiktoren in gleicher Weise mit WM und gf. So hingen FSK stärker mit den kognitiven Faktoren zusammen, als IM. Lediglich im Fach Mathematik fand sich eine hoch signifikante Korrelation von $r = .16$ zwischen IM und WM. Zu gf konnte in beiden Fächern kein signifikanter Zusammenhang festgestellt werden. Im Unterschied hierzu ergaben sich für FSK positive Zusammenhänge in einem Range von $r = .14$ bis $r = .32$. Die Assoziation des FSKs zu WM erwies sich mit Werten von $r = .31$ für Deutsch und $r =$

.32 für Mathematik als stärker im Vergleich zu gf. So korrelierte FSK mit gf zu $r = .14$ in Deutsch und $r = .18$ in Mathematik.

Auch der elterliche SES korrelierte signifikant mit gf ($r = .26$), während die Korrelation zu WM mit einem Wert von $r = .15$ geringer ausfiel. Seitens der non-kognitiven Faktoren zeigte sich, dass IM keine signifikante Korrelation zum SES aufwies, während FSK signifikant korrelierte ($r = .18$). Die Betrachtung der Korrelationen getrennt für beide Elternteile ergab, dass der Bildungsabschluss der Mütter etwas höher mit FSK korrelierte ($r = .14/.16$), verglichen mit den Werten der Väter ($r = .11/.14$). In Deutsch erwies sich die Korrelation zum SES des Vaters als nicht signifikant. Allerdings korrelierte der Abschluss des Vaters signifikant negativ mit IM in Deutsch ($r = -.12$). Die Korrelation erreichte jedoch nur auf dem 5%-Niveau Signifikanz.

4.1.5 Deskriptive Statistiken und Binnenstruktur des Kriteriums Schulerfolg

Tabelle 10 gibt die Verteilungskennwerte des Kriteriums Schulnote wieder, sowohl die der einzelnen Halbjahres- und Jahresnoten im Fach Deutsch und Mathematik, als auch die Werte des jeweiligen fachspezifischen Notendurchschnitts. Die hohen Werte hinsichtlich der internen Konsistenz wiesen bereits auf eine hohe Stabilität der Schulnoten hin ($\alpha_d = .92$; $\alpha_m = .88$). Alle Mittelwerte bewegten sich um Werte von $\bar{x} = 2.5$ bis $\bar{x} = 2.6$, was bedeutete, dass der überwiegende Teil der Schüler die Note *gut* erzielte. Dies spiegelte sich auch in den eher niedrigen Standardabweichungen der Noten wider ($SD = .83$ bis $.98$), wobei eine größere Streuung im Fach Mathematik vorlag als in Deutsch. In beiden Fächern lag eine leichte Rechtsschiefe vor (Schiefe = $.30 - .53$). Da die Kennwerte insgesamt jedoch als unkritisch bezeichnet werden konnten, erfolgte keine weitere Prüfung der Normalverteilungsannahme.

Tabelle 10: Deskriptive Statistiken der Schulnoten

	N	α	Min	Max	M	SD	S	K	K-S
Deutsch		.92							
Jahresnote 3. Klasse	304		1	5	2.55	.85	.30	-.22	.25**
Halbjahresnote 4. Klasse	305		1	5	2.59	.88	.48	-.14	.27**
Notendurchschnitt	306		1	5	2.56	.83	.42	-.12	.21**
Mathematik		.88							
Jahresnote 3. Klasse	304		1	5	2.48	.96	.44	-.04	.24**
Halbjahresnote 4. Klasse	306		1	5	2.63	.98	.53	-.13	.25**
Notendurchschnitt	306		1	5	2.56	.92	.52	-.07	.17**

Anmerkungen: α : Cronbachs Alpha; Min: Minimum; Max: Maximum; M: Mittelwert; SD: Standardabweichung; S: Schiefe; K: Kurtosis; K-S: Kolmogorov-Smirnoff-Test korrigiert nach Lilliefors.

Abbildung 6 stellt die Häufigkeiten der Noten getrennt nach Schulfach und Klassenstufe in Prozentwerten dar. Die Note *sehr gut* wurde im Fach Deutsch in 8.2% bzw. in 6.8% der Fälle vergeben. In Mathematik erreichten 13.8% der Schüler der dritten Klasse die Bestnote und 8.8% in der vierten Klasse. Im Fach Deutsch wurde die Note *gut* in der dritten Klasse von 43.4% der Schüler erzielt, in der vierten Klasse von 45.6% der Schüler. Im Fach Mathematik erreichten 40.5% der Schüler in Klasse 3 bzw. 42.5% in Klasse 4 die Note *gut*. Die Note 3 oder *befriedigend* wurde in der Klassenstufe 3 bei 34.9% der Schüler in Deutsch vergeben und bei 32.2% der Schüler in Mathematik. In der vierten Klassenstufe erzielten 31.5% der Kinder eine 3 in Deutsch und 30.4% der Kinder in Mathematik. Die Note 4 (*ausreichend*) wurden in Deutsch in 12.5% bzw. 14.1% der Fälle vergeben, in Mathematik in 10.5% bzw. 13.7% der Fälle. Eine 5 (*mangelhaft*) trat in Deutsch nur bei 1% bzw. 2% der Schüler auf, in Mathematik bei 3% bzw. 4.6%. Die Note *ungenügend* (6) wurde in keinem Fach vergeben.

Zum besseren Verständnis wurden die Noten in den folgenden Analysen umkodiert, so dass höhere Werte einer besseren Leistung entsprachen.

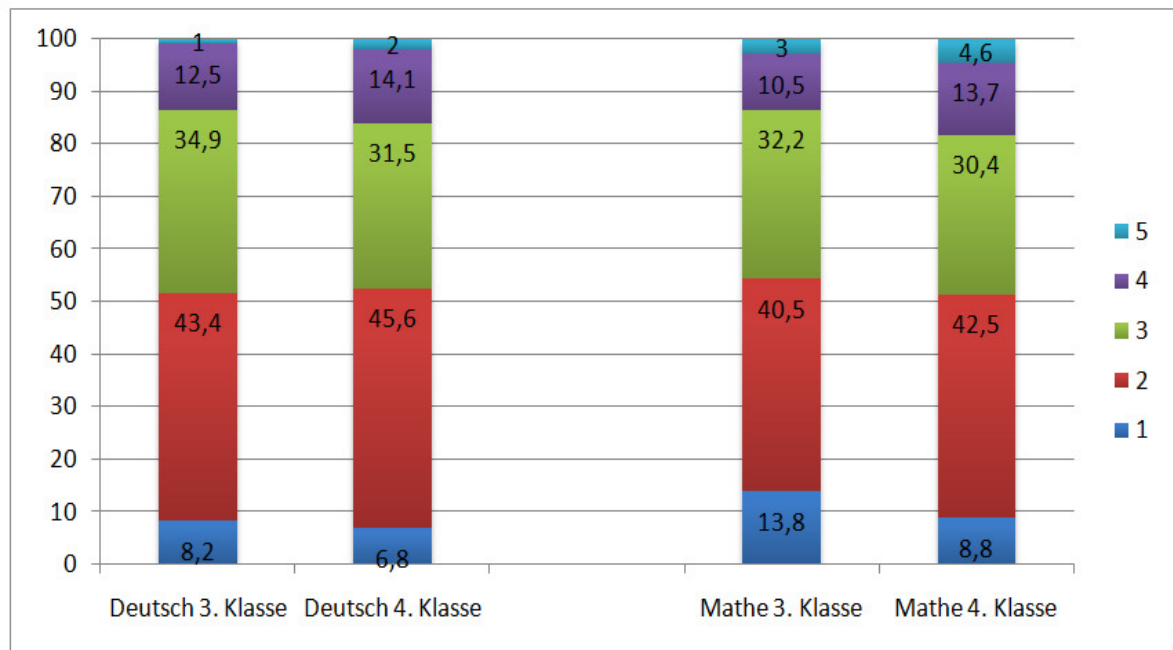


Abbildung 6: Häufigkeitsverteilung der Schulnoten getrennt nach Schuljahr und -fach (Prozentwerte)

Anmerkung: $N \geq 304$; 1 = sehr gut, 2 = gut, 3 = befriedigend, 4 = ausreichend, 5 = ungenügend; Angaben beziehen sich auf gültige Prozente, für welche Notenangaben vorlagen.

Tabelle 11 beinhaltet die Interkorrelationen der Noten in den Fächern Deutsch und Mathematik. Alle Korrelationen erwiesen sich als hoch signifikant und lagen in einem Bereich $r > .65$, weshalb die Zusammenhänge durchweg als hoch zu bezeichnen sind. Konvergente Korrelationen fielen höher aus als divergente. So korrelierten die Deutsch- und Mathematiknoten zwischen $r = .65$ bis $r = .68$, während die Schulnoten innerhalb der Domäne Deutsch zu $r = .85$ korrelierten und im Fach Mathematik zu $r = .78$. Die jeweiligen Durchschnittsleistungen korrelierten mit den einzelnen Schulnoten zu $r = .96$ in Deutsch und zu $r = .94$ in Mathematik. Über die Domänen hinweg fand sich ein Zusammenhang zwischen der durchschnittlichen Leistung in Mathematik und Deutsch von $r = .73$.

Tabelle 11: Korrelationen der Schulnoten und des fachspezifischen Notendurchschnitts

	D4	D	M3	M4	M
Deutsch					
Jahresnote 3. Klasse (D3)	.85	.96	.68	.65	.71
Halbjahresnote 4. Klasse (D4)	-	.96	.66	.68	.71
Notendurchschnitt (D)	-	-	.69	.69	.73
Mathematik					
Jahresnote 3. Klasse (M3)	-	-	-	.78	.94
Halbjahresnote 4. Klasse (M4)	-	-	-	-	.94
Notendurchschnitt (M)	-	-	-	-	-

Anmerkungen: N ≥ 303; alle r signifikant mit $p < .01$.

4.1.6 Korrelative Betrachtung der Prädiktor-Kriteriums-Zusammenhänge

Folgende Tabelle 12 zeigt die korrelativen Zusammenhänge der kognitiven und non-kognitiven Prädiktoren sowie des SES mit den mittleren Schulleistungen in den Fächern Deutsch und Mathematik.

Für beide Fächer ergaben sich hohe Zusammenhänge zwischen Schulnote und WM bzw. gf von nahezu identischer Höhe, wobei im Fach Mathematik insgesamt stärkere Zusammenhänge gefunden wurden ($r_{WM} = .45$; $r_{gf} = .45$) als im Fach Deutsch ($r_{WM} = .36$; $r_{gf} = .34$). Auch auf Subtestebene bestätigte sich der Befund einer stärkeren Beziehung der kognitiven Skalen zur Mathematiknote. Dennoch blieb keine Korrelation zur Deutschnote unterhalb der Signifikanzgrenze.

Innerhalb der Gruppe der non-kognitiven Prädiktoren erwies sich FSK als stärker mit den Schulnoten assoziiert als IM. So korrelierte FSK zu $r = .63$ in Deutsch und in Mathematik zu $r = .56$. Auch die Zusammenhänge des FSK Deutsch zur Mathematiknote bzw. des FSK Mathematik zur Deutschnote überschritten mit Werten von $r = .39$ bzw. $r = .37$ die Signifikanzgrenze, auch wenn diese betragsmäßig niedriger ausfielen, als die konvergenten Korrelationen. IM zeigte fachspezifische Zusammenhänge von $r = .27$ in Deutsch und $r = .31$ in Mathematik. Auch hier fielen die divergenten Korrelationen niedriger aus ($r = .13$), bzw. erwiesen sich im Fach Deutsch als nicht signifikant ($r = .04$).

Insgesamt erreichten die Korrelationen der non-kognitiven Variablen im Fach Deutsch höhere Werte verglichen mit Mathematik. FSK korrelierte stärker mit den entsprechenden Noten als die kognitiven Faktoren.

Auch der höchste Bildungsabschluss der Eltern wies signifikante Korrelationen zu den Schulnoten auf (Deutsch: $r = .32$; Mathe: $r = .32$). Die getrennte Betrachtung beider Elternteile zeigte, dass der mütterliche Bildungsabschluss stärker korrelierte als der Abschluss der Väter (Deutsch: $r_M = .32$, $r_V = .28$; Mathe: $r_M = .33$, $r_V = .27$). Im Vergleich zu den kognitiven und non-kognitiven Individualfaktoren war festzustellen, dass in beiden Fächern der elterliche Bildungsabschluss geringere korrelative Zusammenhänge aufwies.

Tabelle 12: Korrelative Zusammenhänge der Prädiktoren mit dem fachspezifischen Notendurchschnitt

	D	M
Kognitive Prädiktoren		
Arbeitsgedächtnis (WM)	.36**	.45**
Digit Sorting (DS)	.26**	.37**
Animal Sorting (AS)	.37**	.34**
Backward Matrix Span (MS)	.13**	.26**
Intelligenz (gf)	.34**	.45**
CFT 1	.32**	.36**
CFT 2	.20**	.27**
CFT 3	.27**	.35**
CFT 4	.15**	.27**
Non-kognitive Prädiktoren		
Fähigkeitsselbstkonzept Deutsch (FSKd)	.63**	.39**
Fähigkeitsselbstkonzept Mathe (FSKm)	.37**	.56**
Intrinsische Motivation (IMd)	.27**	.04
Intrinsische Motivation (IMm)	.13*	.31**
SES		
Höchster Bildungsabschluss Mutter (SES _M)	.32**	.33**
Höchster Bildungsabschluss Vater (SES _V)	.28**	.27**
Höchster Bildungsabschluss Eltern (SES _E)	.32**	.32**

Anmerkungen: N > 276; D: Notendurchschnitt Deutsch; M: Notendurchschnitt Mathematik;
* = $p < .05$; ** = $p < .01$.

Nachdem hiermit alle vorbereitenden Betrachtungen der Binnenstruktur aller Konstrukte, der Verteilungen sowie der korrelativen Zusammenhänge abgeschlossen sind, folgt nun die Darstellung der weiterführenden Analysen. Zunächst werden die regressions- bzw. kommunalitätenanalytischen Befunde bezüglich der Bedeutung von WM und gf bei der Vorhersage des Schulerfolgs dargestellt, bevor die Strukturgleichungsanalysen berichtet werden.

4.2 Die Bedeutung der kognitiven Faktoren WM und gf als Prädiktoren des Schulerfolgs

4.2.1 Regressionsanalytische Befunde

Mit Bezug auf die in Kapitel 1.2.6 besprochene Diskussion um die Beziehung zwischen Intelligenz und Arbeitsgedächtnis bzw. den Grad der Überschneidung beider Konstrukte, wurden Regressions- bzw. Kommunalitätenanalysen durchgeführt, um auf das Verhältnis geteilter und spezifischer Varianzanteile bei der Vorhersage der Schulnoten rückschließen zu können. Die diesbezüglichen Befunde werden im folgenden Absatz dargelegt.

Auch wenn regressionsanalytische Verfahren gegenüber der latenten Modellierung im Rahmen von Strukturgleichungsanalysen gewisse methodische Nachteile mit sich bringen (siehe Kapitel 3.5.2), muss dennoch an dieser Stelle auf diese Methode zurückgegriffen werden. Die Frage bezüglich des Verhältnisses von gemeinsamer und spezifischer Varianz der beiden kognitiven Prädiktoren ließ sich im Falle der vorliegenden Studie nicht durch Modellierung von Strukturgleichungsanalysen klären, da für den latenten Arbeitsgedächtnisfaktor lediglich drei Indikatoren zur Verfügung standen und eine Aufsplittung in spezifische und mit Intelligenz geteilte Varianzpfade in einer Unteridentifizierung des Modells mündete.

Gemäß Hypothese 1b wurde erwartet, dass beide kognitiven Prädiktoren substantielle Varianzanteile im Kriterium Schulerfolg aufklären. Weiterhin wurde angenommen, dass neben spezifischen Varianzanteilen ein deutlicher Anteil an geteilter Varianz nachgewiesen werden kann (Hypothese 1a).

4.2.1.1 Regressionsanalytische Befunde Deutsch

Die folgende Tabelle 13 stellt die Ergebnisse der Regression der mittleren Schulleistung im Fach Deutsch auf WM und gf dar.

Tabelle 13: Regression der mittleren Leistung in Deutsch auf WM und gf

Block	P	β	T	p	R	R^2	Korr. R^2	ΔR^2	ΔF	Δp (F)
(df1,df2)										
A	1	WM	.26	4.56	.00	.42	.18	.17	.18	32.33
		gf	.23	3.85	.00					<.001 (2,303)
B	1	WM	.37	6.84	.00	.36	.13	.13	.13	46.77
										<.001 (1,304)
	2	WM	.26	4.56	.00	.42	.18	.17	.05	15.64
		gf	.23	3.95	.00					<.001 (1,303)
C	1	gf	.34	6.42	.00	.34	.12	.12	.12	41.19
										<.001 (1,304)
	2	gf	.23	3.95	.00	.42	.18	.17	.06	20.79
		WM	.26	4.56	.00					<.001 (1,303)

Anmerkungen: N=305; Regressand: Mittelwert der beiden Deutschnoten; P: Prädiktor; WM: Arbeitsgedächtnisfaktor, gf: Intelligenzfaktor; Block 1: Methode Einschluss; Block 2: Methode schrittweise.

In Modell A gingen WM und gf gleichzeitig als Prädiktoren in die Regression ein. Beide Variablen erklärten gemeinsam 17.0% der Varianz im Kriterium Deutschnote auf. Gemäß Hypothese 1b wiesen sowohl WM als auch gf signifikante und nahezu identische Beta-Werte auf ($\beta_{WM} = .26$, $\beta_{gf} = .23$). Im Zuge einer hierarchischen Regression wurde anschließend die inkrementelle Validität von gf über WM hinaus überprüft (Modell B). WM erklärte 13.0% der Varianz im Kriterium, bei einem signifikanten Betagewicht von $\beta_{WM} = .37$. Gf leistete mit einer zusätzlichen Varianzaufklärung von 4.3% einen inkrementellen Beitrag über WM hinaus ($\beta_{gf} = .23$). In einem dritten Modell (Modell C) wurde schließlich die inkrementelle Validität von WM über gf hinaus überprüft. Es zeigte sich eine inkrementelle Varianzaufklärung von 5.7%. Intelligenz alleine erklärte 11.9% der Varianz im Kriterium, ein Wert, der der prädiktiven Validität von WM nahezu entsprach.

Um die aufgeklärte Varianz in gemeinsame und spezifische Varianzanteile von gf und WM zu zerlegen, wurde eine Kommunalitätenanalyse durchgeführt. Diese ergab, dass relativiert an der insgesamt aufgeklärten Varianz von 17.0%, WM einen spezifischen Varianzanteil von 31.3% aufwies. 23.6% der Varianz konnten spezifisch auf gf zurückgeführt werden. Ein substantieller Varianzanteil (45.1%) ging auf die durch WM und gf geteilte Varianz zurück. Das Verhältnis von spezifischer und geteilter Varianz ist in Abbildung 7 graphisch dargestellt.

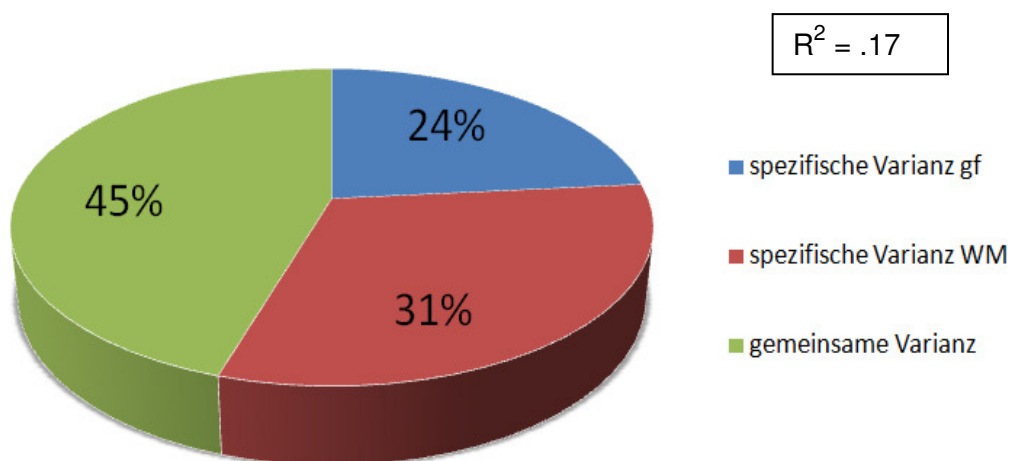


Abbildung 7: Spezifische und gemeinsame Varianzanteile von gf und WM im Fach Deutsch

Im Anschluss an die Analysen für Deutsch wurden analoge Analysen für die mittlere Leistung in Mathematik durchgeführt.

4.2.1.2 Regressionsanalytische Befunde Mathematik

Die Ergebnisse der Regressionsanalysen im Fach Mathematik (Tabelle 14) zeigten, dass mit insgesamt 28.1% aufgeklärter Varianz über 10% mehr Varianz innerhalb des Kriteriums aufgeklärt werden konnten als in Deutsch. In Modell A gingen WM und gf als gleichberechtigte Prädiktoren ein. Es ergaben sich nahezu identische Betagewichte für beide Variablen ($\beta_{WM} = .31$, $\beta_{gf} = .32$). Hinsichtlich der inkrementellen Validität zeigte Modell B, dass gf in Mathematik 8.1% zusätzliche Varianz über WM hinaus aufklären konnte, während WM ein Inkrement von 7.7%

über gf hinaus aufwies (Modell C). WM alleine erzielte eine Varianzaufklärung von 20.0%, gf alleine erklärte 20.4%.

Tabelle 14: Regression der mittleren Leistung in Mathematik auf WM und gf

Block		P	β	T	p	R	R^2	Korr. R^2	ΔR^2	ΔF	Δp (F)
		(df1,df2)									
A	1	WM	.31	5.70	.00	.53	.28	.28	.28	59.14	<.001
		gf	.32	5.82	.00					(2,303)	
B	1	WM	.45	8.72	.00	.45	.20	.20	.20	76.10	<.001
										(1,304)	
	2	WM	.31	5.70	.00	.53	.28	.28	.08	33.93	<.001
		gf	.32	5.82	.00					(1,303)	
C	1	gf	.45	8.81	.00	.45	.20	.20	.20	77.71	<.001
										(1,304)	
	2	gf	.32	5.82	.00	.53	.28	.28	.08	32.51	<.001
		WM	.31	5.70	.00					(1,303)	

Anmerkungen: N=305; Regressand: Mittelwert der beiden Mathematiknoten; P: Prädiktor; WM: Arbeitsgedächtnisfaktor, gf: Intelligenzfaktor; Block 1: Methode Einschluss; Block 2: Methode schrittweise.

Zur Quantifizierung des Verhältnisses von geteilter und ungeteilter Varianz wurde auch hier eine Kommunalitätenanalyse durchgeführt. Relativiert an der insgesamt aufgeklärten Varianz von 28.1%, ergab sich ein spezifischer Varianzanteil von 28.5%, der auf gf zurückzuführen war. 26.5% der Varianz erwiesen sich als spezifisch für WM. Daraus resultierend teilten beide Konstrukte 45.0% der Varianz.

Abbildung 8 illustriert das Verhältnis von spezifischer und gemeinsamer Varianz.

Somit konnten in beiden Fächern sowohl substantielle Anteile an gemeinsamer Varianz nachgewiesen werden, als auch spezifische Varianzanteile. Die prädiktive Validität von WM und gf in Mathematik fiel insgesamt höher aus als in Deutsch.

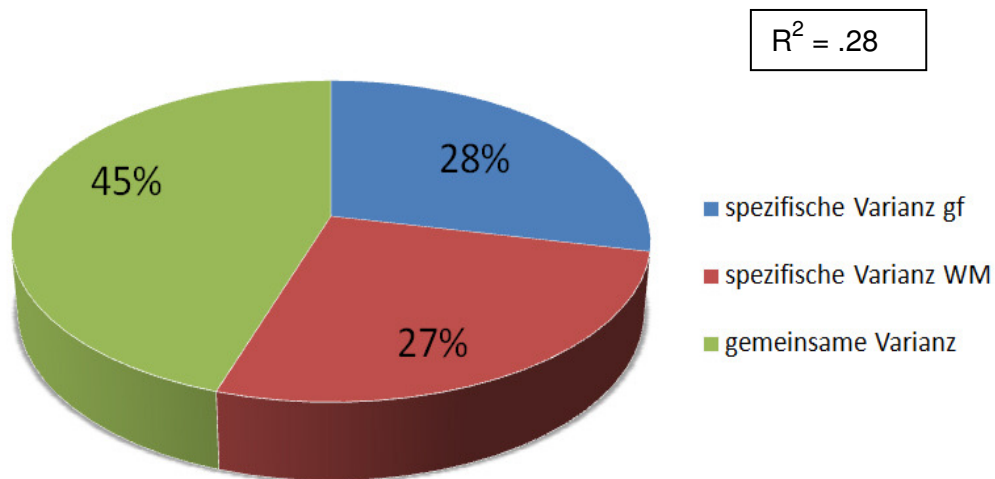


Abbildung 8: Spezifische und gemeinsame Varianzanteile von gf und WM im Fach Mathematik

Nach der Untersuchung der relativen Bedeutung von WM und gf hinsichtlich der Vorhersage des Schulerfolgs auf manifester Ebene, erfolgte die Betrachtung auf latenter Ebene. Diesbezügliche Analysen werden im folgenden Kapitel dargestellt.

4.2.2 Strukturgleichungsanalytische Befunde

4.2.2.1 Voraussetzungsprüfung

Wie in Kapitel 3.5.2 erläutert, sollte vor der eigentlichen Modellprüfung die Prüfung der Voraussetzungen erfolgen. Da die Überprüfung der Skalen auf univariate Normalverteilung bzw. auf univariate Ausreißer bereits erfolgt war (Kapitel 3.5.2), wurde nun die Prüfung auf multivariate Ausreißer durchgeführt, mittels der in Kapitel 3.5.2 beschriebenen Methode zur Berechnung der Mahalanobis Distanz (Tabachnik & Fidell, 2001). Zwei Fälle wurden auf diesem Wege als multivariate Ausreißer identifiziert und in folgenden Analysen exkludiert.

Homoskedastizität und Linearität wurden mittels der Betrachtung der in Kapitel 3.5.2 beschriebenen Streudiagramme geprüft.

Multikollinearität konnte durch Inspektion der Toleranzwerte sowie der Korrelationen zwischen den untersuchten Variablen ausgeschlossen werden, da keine Korrelation

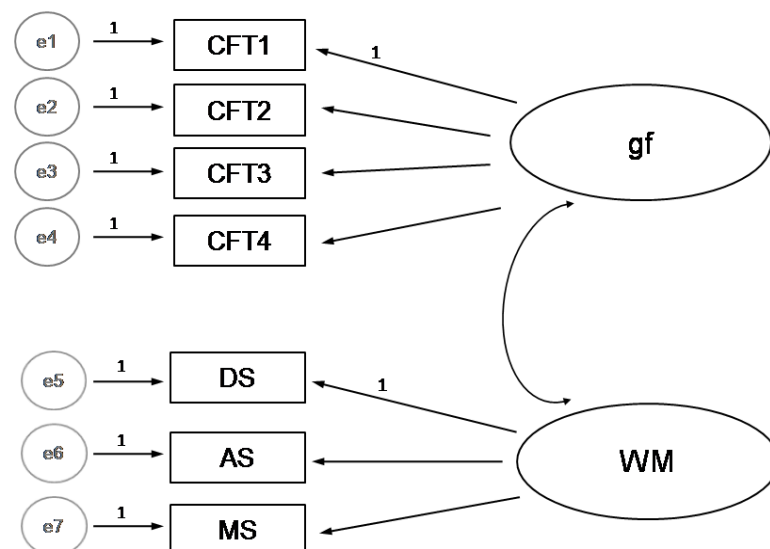
von $r > .85$ vorlag und die Toleranz mit Werten von $Tol > .46$ deutlich oberhalb des kritischen Wertes von $Tol < .10$ lag (Bühner, 2006).

Die Ergebnisse der Mardia-Tests bezüglich multivariater Normalverteilung werden bei den jeweiligen Modellierungen berichtet.

4.2.2.2 Prüfung des Messmodells der kognitiven Faktoren

Bevor das eigentliche Vorhersagemodell der kognitiven Prädiktoren getestet wurde, wurde in einem nächsten Schritt zunächst die Gültigkeit des angenommenen Messmodells überprüft. Der latente Intelligenzfaktor (gf) wurde durch vier Indikatoren repräsentiert (CFT Subtests 1 bis 4). Die drei Arbeitsgedächtnisaufgaben (DS, AS, MS) dienten als Indikatoren des latenten Arbeitsgedächtnisfaktors (WM). Zwischen den latenten Variablen für WM und gf wurde eine Korrelation zugelassen. Pro latentem Faktor wurde ein Pfadgewicht auf eins fixiert, zur Festsetzung der Modellmetrik. Abbildung 9 illustriert das Messmodell.

Abbildung 9: Zweifaktorielles Messmodell der kognitiven Fähigkeiten

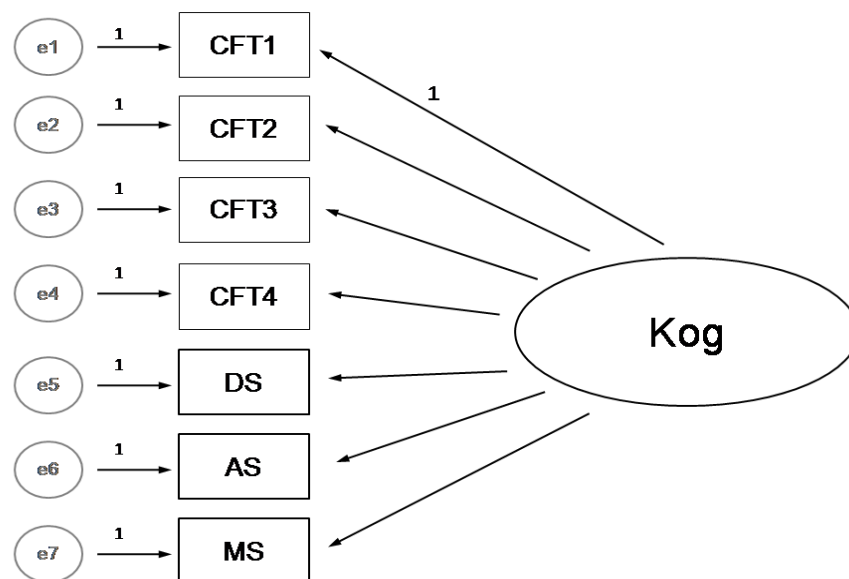


Anmerkungen: CFT1-4: CFT-Subtests; DS, AS, MS: WM-Tasks; gf: latenter Intelligenzfaktor; WM: latenter Arbeitsgedächtnisfaktor; e1-7: Fehlerterme.

Die Annahme der multivariaten Normalverteilung konnte durch den Mardia-Test bestätigt werden (Multivariate Kurtosis = 1.83, c.r. = 1.45). Das Modell wurde mittels der ML- Methode geschätzt, welche eine hervorragende Passung lieferte (CMIN = 13.99, df = 13, p = .37; GFI = .99; CFI = 1; RMSEA = .02 (0 - .06); AIC = 43.99). Alle Faktorladungen der Indikatoren erreichten hoch signifikante Werte (standardisierte Lösung: $\lambda_{DS} = .65$; $\lambda_{AS} = .55$; $\lambda_{MS} = .45$; $\lambda_{CFT1} = .61$; $\lambda_{CFT2} = .46$; $\lambda_{CFT3} = .64$; $\lambda_{CFT4} = .52$). Die signifikante Korrelation von $r = .71$ zwischen den latenten Faktoren kann als hoch bezeichnet werden. Es ist zu beachten, dass für dieses Modell keine fehlenden Werte ersetzt werden mussten, so dass sich das Modell auf den vollständigen Datensatz von N=314 bezieht.

Um abzusichern, dass die getrennte Modellierung zweier latenter Variablen für WM und gf den Daten besser entsprach, als die Schätzung eines gemeinsamen latenten Kognitionsfaktors, wurde ein Ein-Faktor-Modell geschätzt und mit dem Zwei-Faktoren-Modell verglichen (Abbildung 10).

Abbildung 10: Einfaktorielles Messmodell der kognitiven Fähigkeiten



Anmerkungen: CFT1-4: CFT-Subtests; DS, AS, MS: WM-Tasks; Kog: latenter Kognitionsfaktor; e1-7: Fehlerterme.

Das Ein-Faktor-Modell lieferte schlechtere Werte für alle Fit- Indizes (CMIN = 31.74, df = 14, p = .00; GFI = .97; CFI = .94; RMSEA = .06 (.03 - .09); AIC = 59.74). Außerdem erwiesen sich alle Faktorladungen als niedriger, verglichen mit dem Zwei-

Faktoren-Modell (standardisierte Lösung: $\lambda_{DS} = .53$; $\lambda_{AS} = .44$; $\lambda_{MS} = .43$; $\lambda_{CFT1} = .59$; $\lambda_{CFT2} = .44$; $\lambda_{CFT3} = .62$; $\lambda_{CFT4} = .49$), was als Bestätigung der Annahme zweier getrennter Faktoren für WM und gf gewertet werden kann und somit das angenommene zweifaktorielle Messmodell bestätigte, welches in den folgenden Analysen beibehalten wurde.

4.2.2.3 Allgemeines Prognosemodell der kognitiven Prädiktoren und des sozioökonomischen Status

Im Anschluss an die Testung des Messmodells der kognitiven Prädiktoren wurde das Prognosemodell geprüft, in welchem Schulleistung getrennt für Deutsch und Mathematik durch WM und gf vorhergesagt wurde, bei gleichzeitiger Modellierung des SES der Eltern. Die höchsten Bildungsabschlüsse beider Elternteile wurden als Indikatoren des latenten SES-Faktors herangezogen. Der latente Schulleistungsfaktor umfasste zwei Indikatoren, die jeweilige Jahresnote der Klassenstufe 3, sowie die Halbjahresnote der vierten Klassenstufe. Pro latentem Faktor wurde eine Faktorladung auf eins fixiert. Abbildung 11 stellt den Aufbau des vollständigen Prognosemodells in Deutsch dar. Es ist zu beachten, dass sich das Modell für Mathematik lediglich darin unterschied, dass statt der Deutschnoten die Noten für Mathematik als Indikatoren der Schulleistung dienten. Fälle ohne Notenangaben (N = 13) gingen nicht in die Modellierung ein.

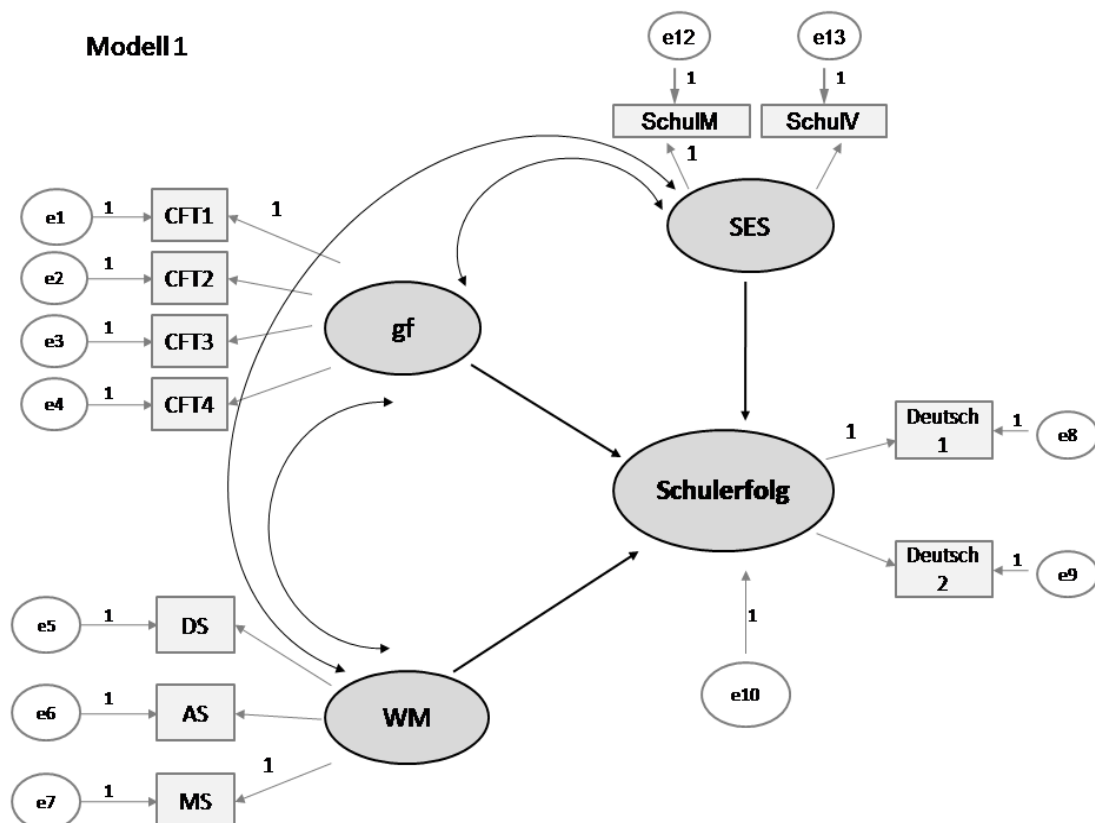
4.2.2.3.1 Untersuchung der prädiktiven Validität von WM, gf und SES in Deutsch

Im Fach Deutsch (Modell 1) zeigte sich ein hervorragender Modellfit (N = 301; CMIN = 46.82, df = 38, p = .15; GFI = .97; CFI = .99; RMSEA = .03 (0 - .05); AIC = 102.82) bei bestätigter Normalverteilungsannahme (Multivariate Kurtosis: 4.29, c.r.: 2.20). Die Pfadgewichte von WM und elterlichem SES zur Schulleistung in Deutsch erwiesen sich als signifikant ($\beta_{WM} = .41$; $\beta_{SES} = .26$). Der Pfad von gf zu Schulerfolg war mit einem $\beta_{gf} = .04$ nicht signifikant. Die Korrelation zwischen WM und gf betrug $r = .69$, zwischen WM und SES $r = .20$ und zwischen SES und gf $r = .45$. Damit lag eine stärkere Assoziation des SES zum latenten gf-Faktor als zum latenten WM-Faktor vor, was dem Ergebnismuster der manifesten Korrelationen entsprach.

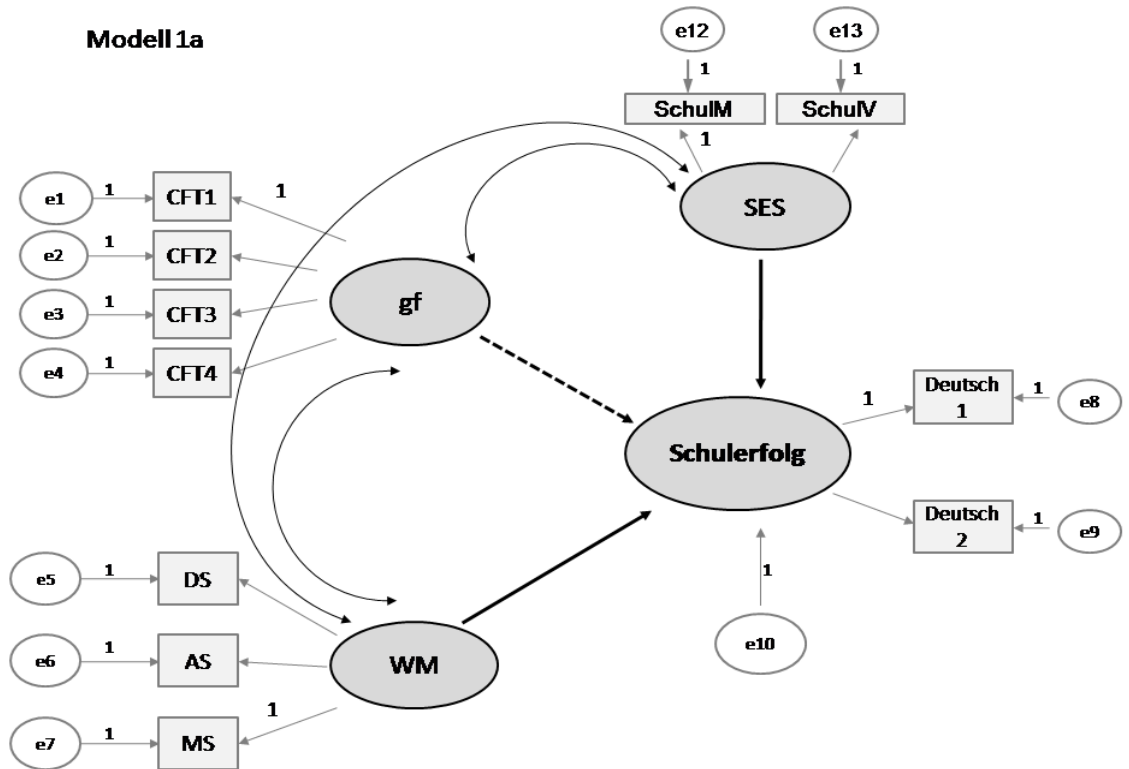
Diese Befunde spiegeln die in Abschnitt 4.2.1 berichteten Ergebnisse der Kommunalitätenanalysen wider, da auch dort der spezifische Varianzanteil von gf an den Deutschnoten mit 24% hinter WM mit 31% spezifischer Varianz zurückfiel. Der in der Kommunalitätenanalyse gefundene substantielle gemeinsame Varianzanteil im Fach Deutsch (45%) fand sich in der latenten Modellierung in der hohen latenten Korrelation von $r = .69$ wieder. Insgesamt wurden 32% der Varianz der Deutschnoten durch die beiden kognitiven Prädiktoren und SES erklärt.

Im Anschluss an das frei geschätzte Modell 1 wurden verschiedene genestete Modelle getestet. In Modell 1a wurde der nicht signifikante Pfad von gf auf Schulerfolg auf null gesetzt, um ein parsimonischeres Modell zu prüfen. In Modell 1b wurde neben dem Pfad von gf auf Schulerfolg auch die Kovarianz zwischen WM und SES auf null gesetzt (siehe Abbildung 11).

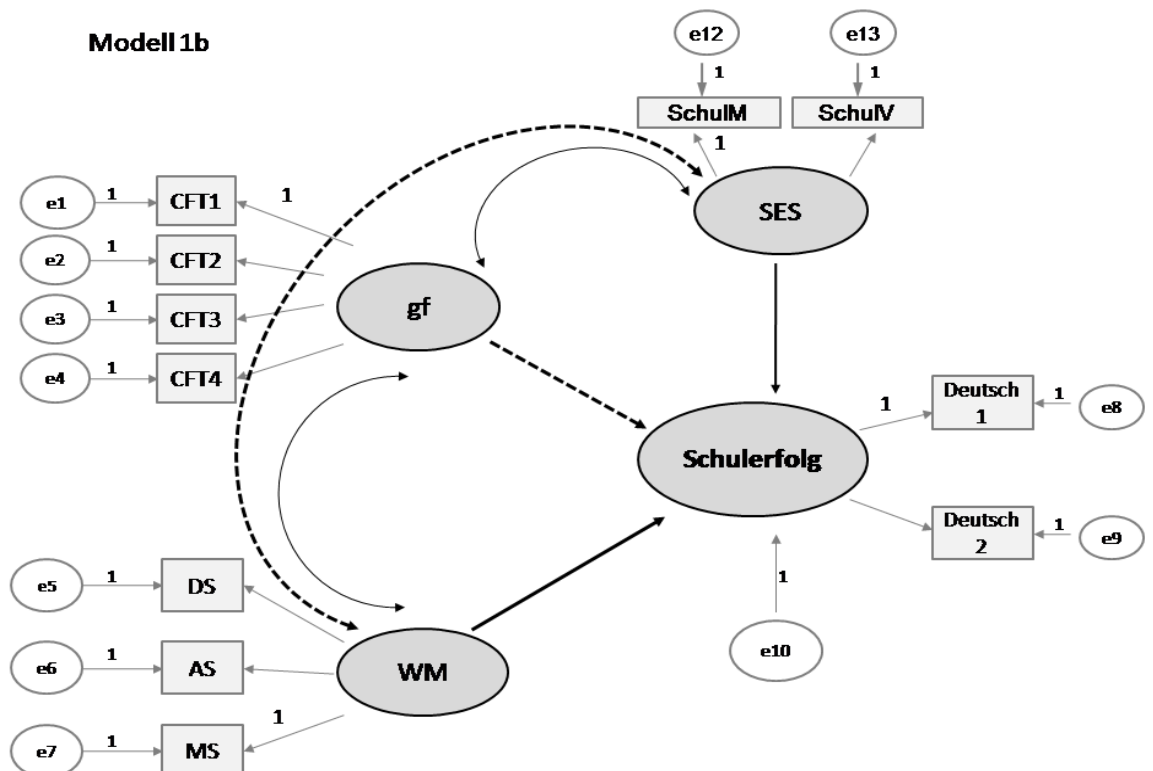
Abbildung 11: Schematische Darstellung der schrittweisen Fixierung nicht signifikanter Pfade in Deutsch



Modell 1a



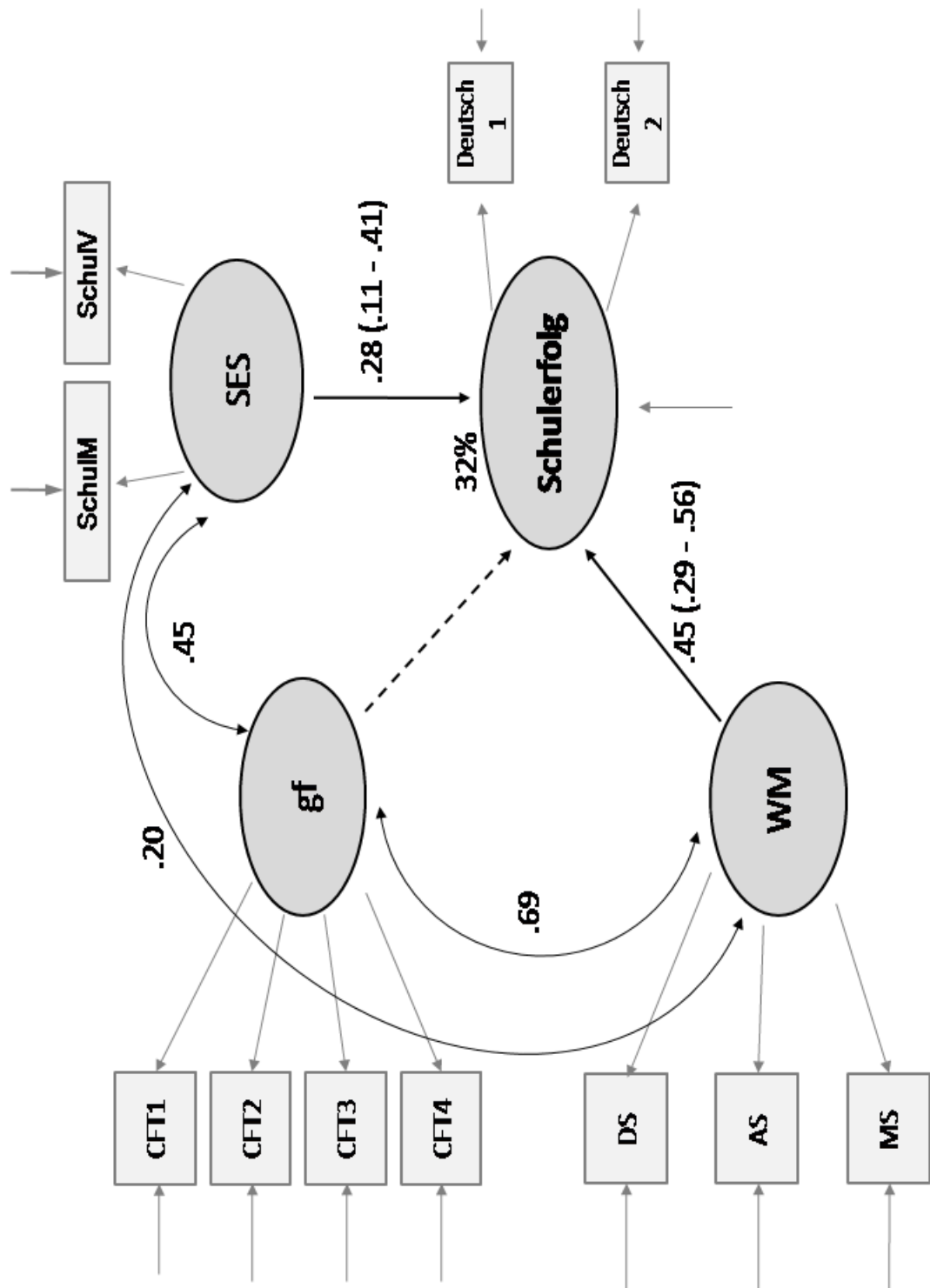
Modell 1b



Anmerkungen: Durchgezogene Linien: signifikante Pfade und Korrelationen;
Gestrichelte Linien: auf null fixierte Pfade und Korrelationen.

Der Modellvergleich ergab, dass lediglich Modell 1a nicht signifikant schlechter fittete als das vollständige Modell, denn Modell 1b lag mit einem Wert von $p = .05$ genau auf der Signifikanzgrenze (siehe Tabelle 15). Vor diesem Hintergrund war Modell 1a aufgrund des Parsimonitätskriteriums dem vollständigen Modell vorzuziehen. Hier stellte WM mit einem standardisierten Pfadgewicht von $\beta_{WM} = .45$ (.29 - .56) den stärksten Prädiktor dar, gefolgt vom elterlichen Bildungsabschluss mit $\beta_{SES} = .28$ (.11 - .41). Die korrelativen Zusammenhänge erwiesen sich als identisch zum vollständigen Modell 1. Insgesamt konnten 32% der Varianz im Kriterium aufgeklärt werden (siehe Abbildung 12).

Abbildung 12: Parsimonischstes Prognosemodell der kognitiven Faktoren und des SES in Deutsch (Modell 1a)



Anmerkungen: N = 301; durchgezogene Linien: signifikante Pfade und Korrelationen; gestrichelte Linien: nicht signifikante Pfade und Korrelationen; Werte in Klammern beziehen sich auf das 95%-Bootstrap-Konfidenzintervall. Aus Gründen der Übersichtlichkeit wurden die Fehlerterme nur angedeutet.

4.2.2.3.2 Untersuchung der prädiktiven Validität von WM, gf und SES in Mathematik

Die Schätzung des vollständigen Prognosemodells in Mathematik (Modell 2) wurde analog zum in Abbildung 11 dargestellten Modell durchgeführt. Im Fach Mathematik änderte sich das Ergebnismuster dahingehend, dass sich die standardisierten Pfadgewichte von gf und SES stärker angleichen ($\beta_{gf} = .23$, $\beta_{SES} = .21$). Obwohl gf hier ein betragsmäßig höheres Betagewicht erreichte als SES lag der Wert für gf aufgrund eines größeren Schätzfehlers unterhalb der Signifikanzgrenze.

WM erwies sich auch in Mathematik mit einem Gewicht von $\beta_{WM} = .41$ als stärkster Prädiktor. Die Korrelation zwischen beiden latenten kognitiven Variablen betrug $r = .69$. Zwischen SES und gf wurde ein Wert von $r = .45$ und zwischen WM und SES von $r = .20$ erreicht. Damit änderten sich die Zusammenhänge zwischen den Prädiktoren nicht im Vergleich zum Fach Deutsch. In Bezug auf die Befunde der Kommunalitätenanalyse war festzustellen, dass sich zwar auch in dieser Domäne der Anteil an geteilter Varianz der beiden kognitiven Variablen in Form einer hohen latenten Korrelation widerspiegelte. Durch die Hinzunahme des SES als weiteren Prädiktor schränkte sich jedoch der Anteil spezifischer gf-Varianz weiter ein, da gf sowohl mit WM als auch mit SES substantiell korrelierte.

Insgesamt konnten 48% der Varianz der Mathematiknoten aufgeklärt werden, 16% mehr als im Fach Deutsch. Auch diese Modellschätzung erzielte eine hervorragende Güte für das vollständige Modell (CMIN = 37.14, df = 38, $p = .51$; GFI = .98; CFI = 1; RMSEA = 0 (0 - .04); AIC = 93.14) bei bestätigter Normalverteilungsannahme (multivariate Kurtosis = 2.23; $p = 1.14$).

Analog zur Vorgehensweise im Fach Deutsch wurden im Anschluss an die Schätzung des vollständigen Modells verschiedene genestete Modelle geprüft. In Modell 2a wurde der nicht signifikante Pfad von gf zu Schulerfolg auf null gesetzt. In Modell 2b¹¹ wurden sowohl der Pfad von gf zu Schulerfolg, als auch die Kovarianz zwischen SES und WM fixiert. Der Modellvergleich ergab, dass Modell 2a nicht signifikant schlechter fittete als das Gesamtmodell, weshalb diesem der Vorzug zu geben war (siehe Tabelle 15). WM erwies sich in diesem Modell mit einem

¹¹ Die Modellierung der genesteten Modelle entsprach dem Vorgehen in Deutsch, daher kann zur Veranschaulichung Abbildung 11 herangezogen werden.

standardisiertes Pfadgewicht von $\beta_{WM} = .60$ (.44 - .72) als stärker, verglichen mit dem elterlichen SES ($\beta_{SES} = .28$ (.11 - .42)) (siehe Abbildung 13). Die Konfidenzintervalle beider Faktoren überlappten nicht. Die Korrelation zwischen den Prädiktoren veränderten sich im Vergleich zu Modell 2 nicht, mit Ausnahme der Korrelation zwischen gf und WM, welche auf den Wert $r = .74$ anstieg. Insgesamt bestätigten sich die regressionsanalytischen Befunde hinsichtlich der stärkeren prädiktiven Validität kognitiver Variablen für das Fach Mathematik. Es wurden 50% der Kriteriumsvarianz durch Modell 2a aufgeklärt.

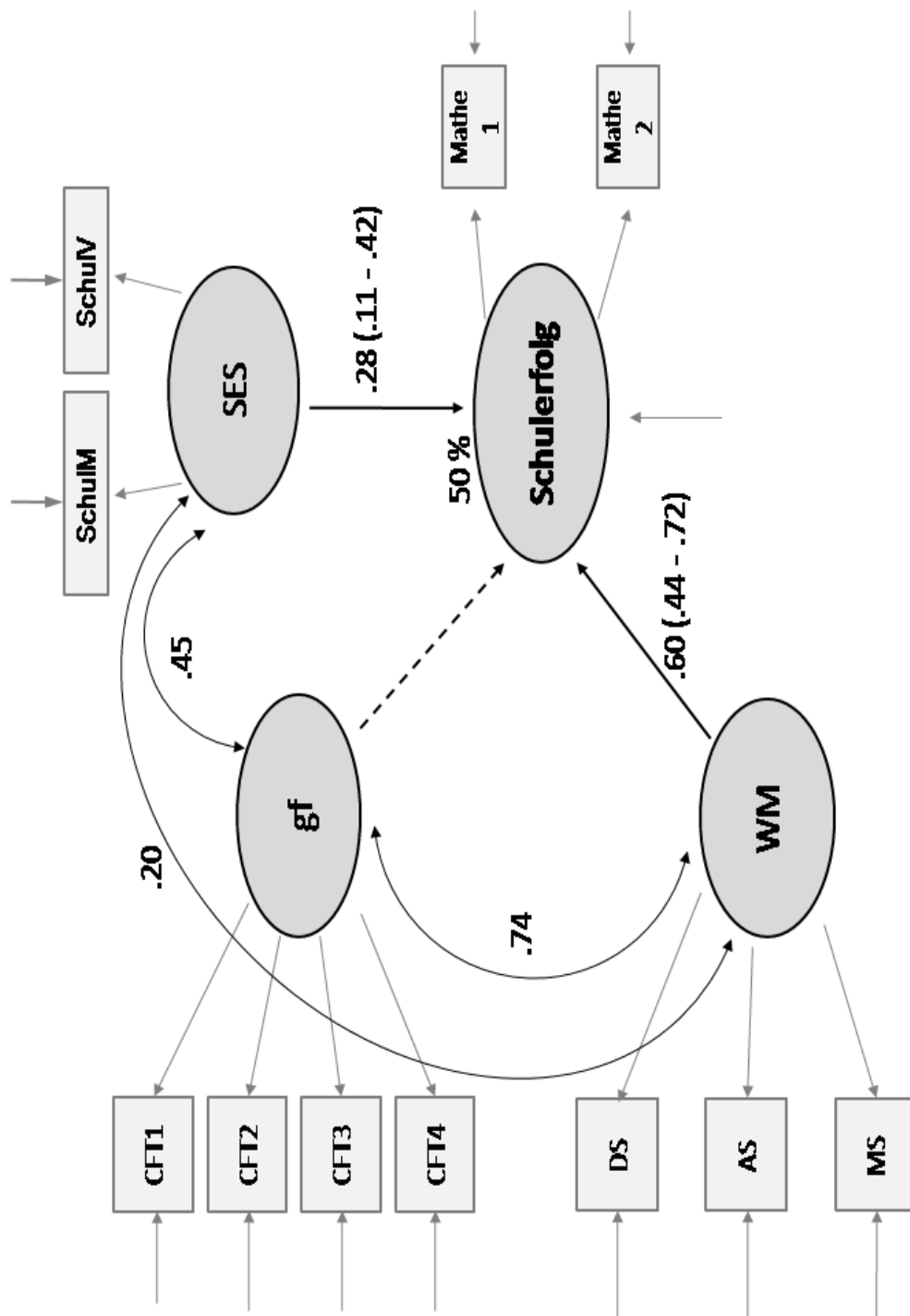
Tabelle 15 fasst die Fit-Indizes der Modelle in Deutsch und Mathematik zur Veranschaulichung zusammen.

Tabelle 15: Fit-Indizes der Prognosemodelle der kognitiven Prädiktoren und des SES in Deutsch und Mathematik

	CMIM	df	p	GFI	CFI	RMSEA	AIC	CMIN*(df,p)
Deutsch								
Modell 1	46.82	38	.15	.97	.99	.03 (0 - .05)	102.82	
Modell 1a	46.89	39	.78	.97	.99	.03 (0 - .05)	100.89	0.07 (1,.79)
Modell 1b	52.79	40	.08	.97	.99	.03 (0 - .05)	104.79	5.97 (2,.05)
Mathematik								
Modell 2	37.14	38	.51	.98	1	.00 (0- .04)	93.14	
Modell 2a	39.18	39	.46	.98	1	.00 (0- .04)	93.18	2.04 (1,.15)
Modell 2b	45.21	40	.26	.97	.99	.02 (0 - .05)	97.21	8.07 (2,.02)

Anmerkungen: N = 301; Modell 1/2: Vollständig; Modell 1/2a: gf constrained; Modell 1/2b: gf und Kovarianz zwischen WM und SES constrained; CMIN*: Wert des χ^2 -Differenztests zum Vergleich genesteter Modelle.

Abbildung 13: Parsimonischstes Prognosemodell der kognitiven Faktoren und des SES in Mathematik (Modell 2a)



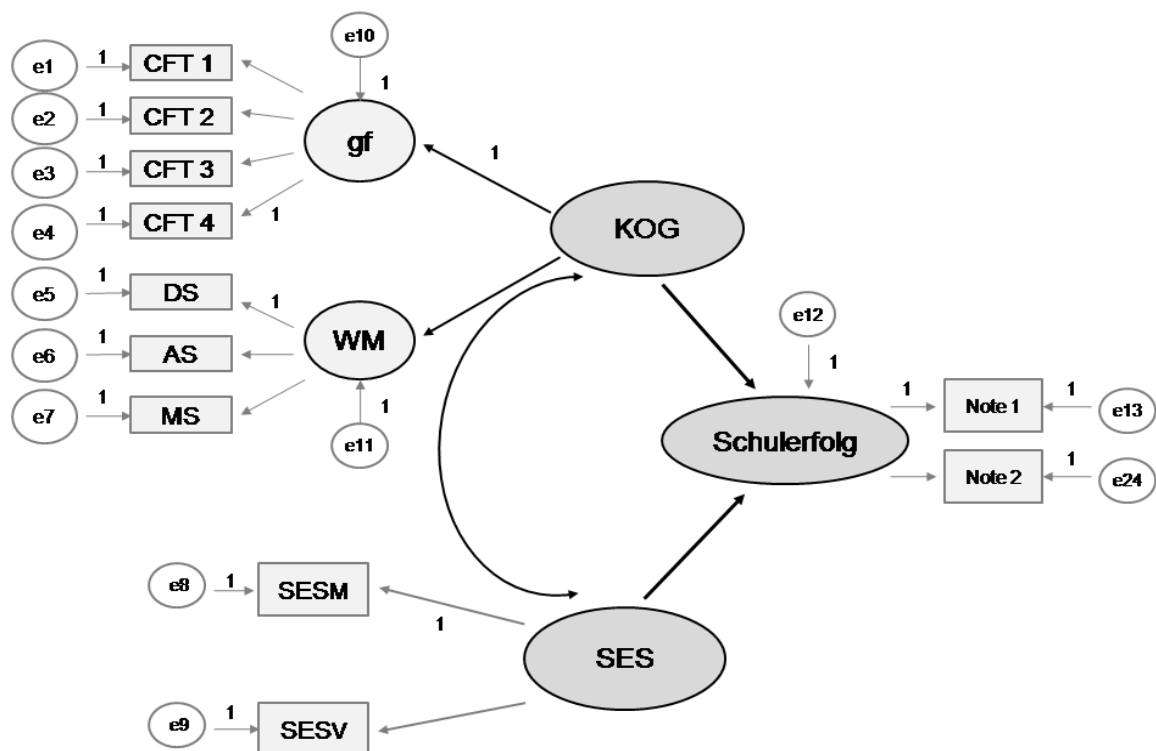
Anmerkungen: N = 301; durchgezogene Linien: signifikante Pfade und Korrelationen; gestrichelte Linien: nicht signifikante Pfade und Korrelationen; Werte in Klammern beziehen sich auf das 95%-Bootstrap-Konfidenzintervall. Aus Gründen der Übersichtlichkeit wurden die Fehlerterme nur angedeutet.

4.2.2.4 Zwei-Ebenen-Modell der kognitiven Prädiktoren und des sozioökonomischen Status in Deutsch

In den zuvor dargestellten Modellen ergab sich das Muster, dass die Variable *gf*, welche Varianzanteile mit der latenten Arbeitsgedächtnisvariable teilte und auch mit der latenten SES-Variable assoziiert war, keine spezifische Varianz im signifikanten Bereich zur Kriteriumsprädiktion beitragen konnte, obwohl das standardisierte Pfadgewicht im Fach Mathematik betragsmäßig höher ausfiel, als das Betagewicht des SES und sich im Fach Deutsch zumindest eine teilweise Überlappung der Konfidenzintervalle fand. Aus den Kommunalitätenanalysen konnte jedoch geschlossen werden, dass WM und *gf* einen beachtlichen Anteil an gemeinsamer Varianz aufweisen. Um die Bedeutung der kognitiven Variablen gegenüber dem elterlichen SES für die Vorhersage der Schulnoten direkter vergleichen zu können, wurde deshalb ein Zwei-Ebenen-Modell modelliert, in welchem die gemeinsame Varianz von *gf* und WM in einem Faktor zweiter Ordnung gebündelt wurde.

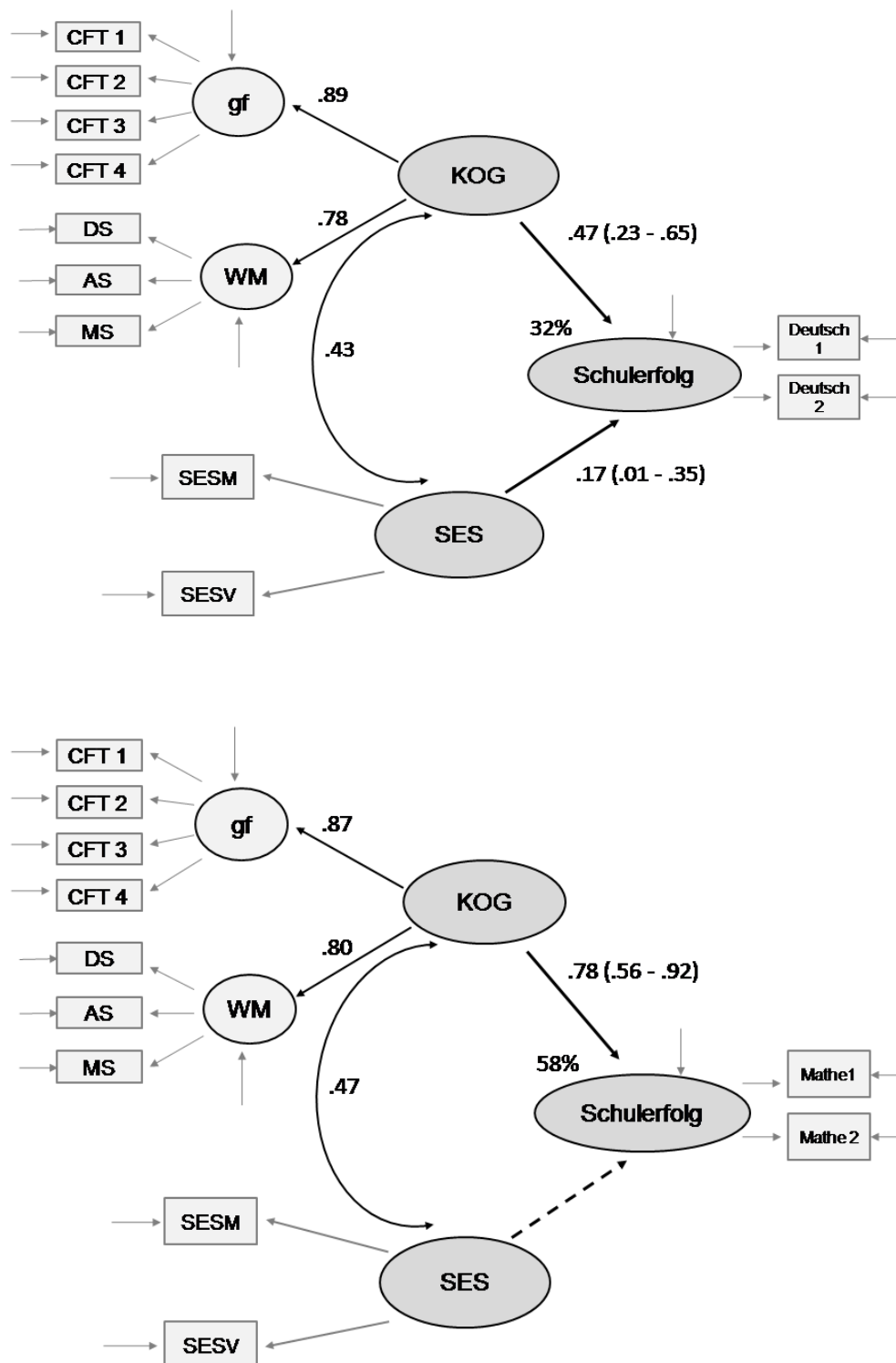
Es wurde ein latenter Kognitionsfaktor (KOG) auf zweiter Ebene den beiden latenten Faktoren erster Ordnung (WM und *gf*) übergeordnet. Der Faktor zweiter Ordnung sowie der latente SES-Faktor dienten der Vorhersage der Schulnote. Es wurde eine Korrelation zwischen dem Kognitionsfaktor zweiter Ordnung und dem latenten SES-Faktor zugelassen (siehe Abbildung 14). Zunächst wurde ein freies Modell ohne Restriktionen geschätzt, welches im Anschluss mit einem parsimonischeren Alternativmodell hinsichtlich des Modellfits verglichen wurde.

Abbildung 14: Zwei-Ebenen-Modell der kognitiven Fähigkeiten und des SES



Der Mardia-Test ergab im Fach Deutsch eine multivariate Kurtosis von 4.29 bei einem c.r.-Wert von 2.20. Die Modellgüte des Zwei-Ebenen-Modells (Modell 3) konnte als gut bezeichnet werden (CMIN: 57.78; df = 39; $p = .03$; GFI = .97; CFI = .98; RMSEA = .04 (.01-.06); AIC = 111.79) bei einer Varianzaufklärung im Kriterium von 32%. Sowohl der latente gf-Faktor als auch die latente WM-Variable waren hoch mit dem allgemeinen Kognitionsfaktor zweiter Ordnung assoziiert, wobei der Wert für gf den Wert für WM überstieg ($\beta_{WM} = .78$; $\beta_{gf} = .89$). Hinsichtlich der prädiktiven Bedeutung erwies sich der Kognitionsfaktor als stärkster Prädiktor der Deutschnote mit einem standardisierten Pfadgewicht von $\beta_{KOG} = .47$ (.23 - .65). Der latente SES-Faktor konnte mit einem $\beta_{SES} = .17$ (.01 - .35) signifikant zur Vorhersage beitragen, wenngleich der Beitrag verglichen mit dem Einfluss des Kognitionsfaktors niedriger ausfiel. Es zeigte sich eine signifikante Assoziation des latenten SES-Faktors zum latenten Kognitionsfaktor ($r = .43$). Die Tatsache, dass das Konfidenzintervall des Pfades von SES zu Schulerfolg nur knapp nicht die null einschloss, gab Anlass zur Überprüfung eines genesteten Modells mit auf null gesetztem Pfad von SES zu Schulerfolg (Modell 3a). Dieses fittete signifikant schlechter (CMIN:4.83; $p = .03$), so dass das vollständige Modell beibehalten wurde (siehe Abbildung 15).

Abbildung 15: Parsimonischstes Zwei-Ebenen-Modell der kognitiven Faktoren und des SES in Deutsch und Mathematik (Modell 3 und Modell 4a)



Anmerkungen: N = 301; durchgezogene Linien: signifikante Pfade und Korrelationen; gestrichelte Linien: nicht signifikante Pfade und Korrelationen; Werte in Klammern beziehen sich auf das 95%-Bootstrap-Konfidenzintervall. Aus Gründen der Übersichtlichkeit wurden die Fehlerterme nur angedeutet.

4.2.2.5 Zwei- Ebenen Modell der kognitiven Prädiktoren und des sozioökonomischen Status in Mathematik

Auch im Fach Mathematik konnte die multivariate Normalverteilungsannahme durch den Mardia-Test bestätigt werden (Multivariate Kurtosis: 2.23; c.r.-Wert: 1.14). Der Modellfit erwies sich als gut (Modell 4) (CMIN: 46.94, df = 39, p = .18; GFI = .97; CFI = .99; RMSEA = .03 (0-.05); AIC=100.94). Die Beziehung des Kognitionsfaktors zweiter Ordnung zu gf und WM änderte sich wenig ($\beta_{WM} = .80$; $\beta_{gf} = .87$), auch die Beziehung des latenten SES-Faktors zu WM und gf blieb stabil ($r = .42$). Unterschiede ergaben sich bezüglich der prädiktiven Validität des latenten Kognitionsfaktors bzw. des latenten SES-Faktors. SES fiel mit einem standardisierten Pfadgewicht von $\beta_{SES} = .12$ stärker hinter die Vorhersagekraft des Kognitionsfaktors zurück ($\beta_{KOG} = .68$). Insgesamt wurden 54% der Kriteriumsvarianz vorhergesagt.

Aufgrund der Tatsache, dass das Konfidenzintervall des Pfadgewichts des SES-Faktors den Wert null einschloss, erfolgte die Testung eines genesteten Modells mit auf null fixiertem Pfad von SES zu Schulerfolg (Modell 4a). Im Gegensatz zum Fach Deutsch fittete dieses parsimonischere Modell nicht signifikant schlechter, weshalb es dem vollständigen Modell vorzuziehen war (CMIN: 2.16; p = .14) (Abbildung 15). Das standardisierte Pfadgewicht des Kognitionsfaktors erreichte einen Wert von $\beta_{KOG} = .78$ (.56 - .92), die Korrelation zwischen KOG und SES belief sich auf $r = .47$. Insgesamt erklärte dieses Modell 58% der Kriteriumsvarianz.

Tabelle 16 veranschaulicht die Fit-Indizes der Zwei-Ebenen-Modelle in Deutsch und Mathematik.

Tabelle 16: Fit-Indizes der Zwei-Ebenen-Modelle der kognitiven Prädiktoren und des SES in Deutsch und Mathematik

	CMIN	df	p	GFI	CFI	RMSEA	AIC	CMIN*(df,p)
Deutsch								
Modell 3	57.78	39	.03	.97	.98	.04 (.01 - .06)	111.79	
Modell 3a	62.83	40	.01	.96	.98	.04 (.02 - .06)	114.62	4.83 (1,.03)
Mathematik								
Modell 4	46.94	39	.18	.97	.99	.03 (.00 - .05)	100.94	
Modell 4a	49.11	40	.15	.97	.99	.03 (.00 - .05)	101.11	2.16 (1,.14)

Anmerkungen: N = 301; Modell 3/4: Vollständig; Modell 3a/4a: SES constrained; CMIN*: Wert des χ^2 -Differenztests zum Vergleich genesteter Modelle.

Zusammengefasst sprechen die in diesem Kapitel dargestellten Modelle, die den Einfluss der kognitiven Prädiktoren und des SES untersuchten, für die Relevanz der kognitiven Prädiktoren Arbeitsgedächtnisses und Intelligenz über die Domänen hinweg. Dabei ist wichtig, die gemeinsamen Varianzanteile zu berücksichtigen (siehe Zwei-Ebenen-Modelle), da die Überlappung beider Konstrukte groß genug war, dass bei einer ausschließlichen Betrachtung spezifischer Varianzanteile ggf. unterhalb der Signifikanzgrenze blieb.

Die fachspezifische Betrachtung zeigte die stärkere Assoziation der kognitiven Faktoren zur mathematischen Domäne. Außerdem fanden sich fachspezifische Unterschiede bezüglich des Einflusses des SES, welcher in der sprachlichen Domäne einen Einfluss ausüben konnte, in Mathematik jedoch an Relevanz verlor. Im Zwei-Ebenen-Modell Mathematik konnte der SES keinen signifikanten Vorhersagebeitrag leisten.

Zur Veranschaulichung der Befunde stellt Tabelle 17 die standardisierten Pfadgewichte und latenten Korrelationen der jeweils parsimonischsten Prognosemodelle der kognitiven Faktoren und des SES in Deutsch und Mathematik dar (Modelle 1a, 2a, 3 und 4a). Die Faktorladungen der manifesten Indikatorvariablen sind Anhang A zu entnehmen.

Tabelle 17: Standardisierte Pfadgewichte und latente Korrelationen der Modelle mit den Prädiktoren WM, gf und SES getrennt nach Schulfach

			Deutsch		Mathe	
			β	r	β	r
Prognosemodell						
gf	→	SE	0	-	0	-
WM	→	SE	.45 (.29 - .56)	-	.60 (.44 - .72)	-
SES	→	SE	.28 (.11 - .41)	-	.28 (.11 - .42)	-
WM	↔	gf	-	.69	-	.74
WM	↔	SES	-	.20	-	.20
gf	↔	SES	-	.45	-	.45
Zwei-Ebenen-Modell						
KOG	→	SE	.47 (.23 - .65)	-	.76 (.65 - .86)	-
SES	→	SE	.17 (.01 - .35)	-	0	-
KOG	→	WM	.78 (.51 - .95)	-	.78 (.56 - .92)	-
KOG	→	gf	.89 (.55 - 1)	-	.86 (.70 - .97)	-
KOG	↔	SES	-	.43	-	.47

Anmerkung: β : standardisiertes Pfadgewicht; r: latente Korrelation; gf: latenter Intelligenzfaktor; WM: latenter Arbeitsgedächtnisfaktor; SES: latenter Faktor des sozioökonomischen Status; KOG: latenter Kognitionsfaktor zweiter Ordnung; SE: Schulerfolg; Werte beziehen sich auf das jeweils parsimonischste Modell getrennt nach Schulfach.

Der Vergleich der Fit-Indizes der ursprünglichen Prognosemodelle mit den Werten der Zwei-Ebenen-Modelle ergab in beiden Fächern den Befund, dass letzere einen marginal schlechteren Fit aufwiesen, der jedoch weiter in einem akzeptablen bis guten Bereich lag. Liegt das Augenmerk auf der Bündelung der gemeinsamen Varianz der kognitiven Variablen, kann das Zwei-Ebenen-Modell daher als geeignet bezeichnet werden.

In der vorliegenden Arbeit sollte jedoch nicht nur die prädiktive Validität der kognitiven Prädiktoren WM und gf betrachtet werden. Ein weiteres Ziel der Arbeit lag im Vergleich der prädiktiven Validität von kognitiven und non-kognitiven Variablen. Deshalb folgt im Anschluss die Darstellung der Modellierungen, welche beide

Prädiktorgruppen berücksichtigten, bei gleichzeitiger Schätzung des Einflusses des elterlichen SES.

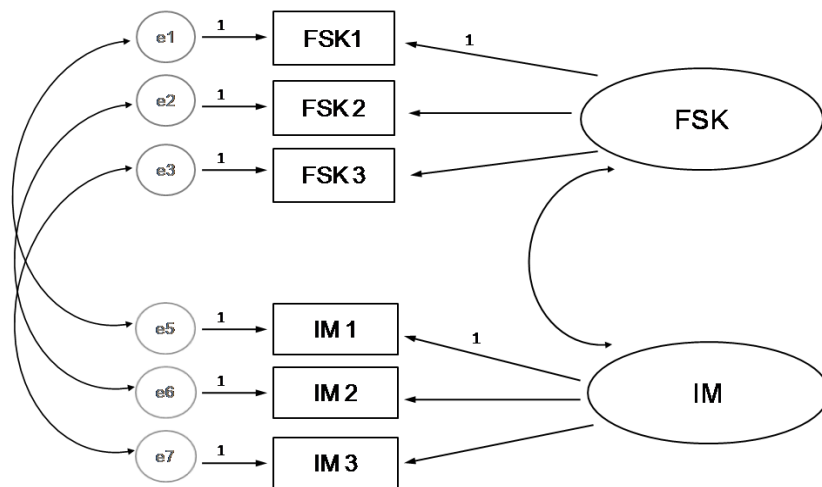
4.3 Relative Bedeutung kognitiver und non-kognitiver Faktoren als Prädiktoren des Schulerfolgs

4.3.1 Prüfung des Messmodells der non-kognitiven Faktoren

Analog zur Vorgehensweise in Kapitel 4.2 wurde zunächst das Messmodell der non-kognitiven Prädiktoren FSK und IM geprüft. Die drei fachspezifischen FSK-Items repräsentierten den latenten FSK-Faktor. Die drei fachspezifischen Items für IM waren mit einem latenten IM-Faktor assoziiert. Zwischen den beiden latenten Faktoren wurde eine Korrelation zugelassen. Abbildung 16 illustriert dieses zweifaktorielle Messmodell.

Von den für die kognitiven Schätzungen verfügbaren Daten von $N = 301$ Personen gingen 23 Fälle nicht in die folgende Schätzung ein, weil keine Fragebogendaten vorlagen. Die innerhalb eines Fragebogens fehlenden Werte für FSK und IM wurden mit der in Kapitel 3.5.2 thematisierten EM-Schätzung ersetzt, so dass sich die folgende Modellschätzung auf einen Datensatz von $N=278$ bezog. Die Missings innerhalb der IM-Skalen erwiesen sich als MCAR (MCAR-Test nach Little: $\chi^2 = 11.86$, $df = 18$, $p = .85$), die Missings der Skala FSK als MAR.

Abbildung 16: Zweifaktorielles Messmodell der non-kognitiven Faktoren



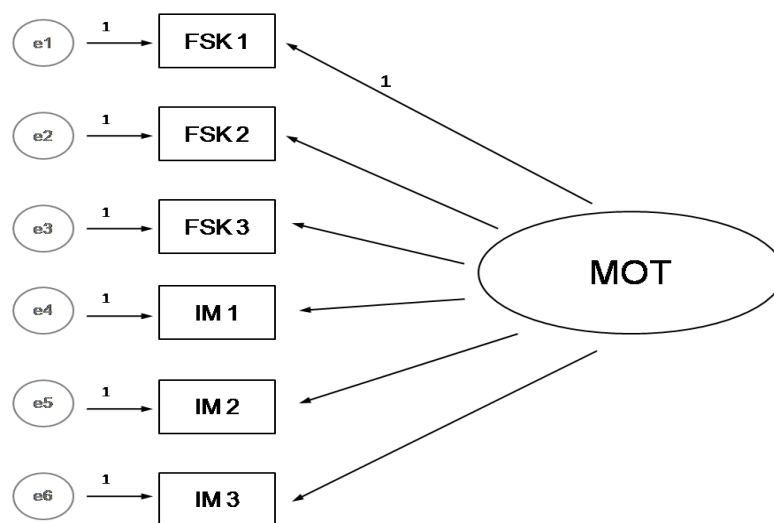
Anmerkungen: FSK1-3: Items zur Erfassung des Fähigkeitsselbstkonzepts; IM1-3: Items zur Erfassung der intrinsischen Motivation; FSK: latenter Faktor Fähigkeitsselbstkonzept; IM: latenter Faktor intrinsische Motivation; e1-7: Fehlerterme.

Das Messmodell fittete für beide Fächer sehr gut (Deutsch: CMIN = 4.80, df = 5, $p = .44$; GFI = .99; CFI = 1; RMSEA = 0 (0 - .08); AIC = 36.80; Mathematik: CMIN = 4.12, df = 5, $p = .53$; GFI = .99; CFI = 1; RMSEA = 0 (0 - .09); AIC = 36.12). In Deutsch variierten die standardisierten Ladungen des latenten FSK-Faktors auf die korrespondierenden Indikatorvariablen zwischen $\lambda_{FSK} = .54$ und $.72$, die Ladungen des IM-Faktors zwischen $\lambda_{IM} = .57$ und $.61$. In Mathematik lagen die Werte der standardisierten Lösung zwischen $\lambda_{FSK} = .64$ bis $.84$ für FSK und zwischen $\lambda_{IM} = .50$ und $.85$ für IM. Somit überschritten alle Werte die Mindestgrenze von $\lambda = .30$, die als Voraussetzung einer sinnvollen Dateninterpretation gilt (Kline, 2005). Die latenten Faktoren korrelierten zu $r = .57$ in Deutsch und zu $r = .74$ in Mathematik. Zu beachten ist, dass Korrelationen zwischen den Fehlern der FSK-Items und den korrespondierenden IM-Items zugelassen wurden, was aufgrund der parallelen Itemformulierung als gerechtfertigt bezeichnet werden konnte. Die Fehlerkorrelationen beeinflussten die Faktorladungen nur gering. Die Korrelation zwischen den beiden latenten Faktoren senkte sich durch das Hinzufügen der Fehlerkorrelationen. Außerdem verbesserte sich der Modell-Fit. Während das Modell ohne korrelierte Fehler aufgrund des schlechten Fits abgelehnt worden wäre (Deutsch: CMIN = 121.61, df = 8, $p = 0$; GFI = .88; CFI = .71; Mathe: CMIN = 128.44, df = 8, $p = .00$; GFI = .88; CFI = .84), zeigte das Modell mit korrelierten Fehlertermen

einen hervorragenden Fit (siehe oben). So legitimierte sich die Modellierung von Korrelationen zwischen Fehlertermen sowohl inhaltlich als auch methodisch (Kline, 2005). Die Gegenüberstellung der Faktorladungen und Korrelationen des Messmodells mit und ohne korrelierte Fehler ist Anhang A zu entnehmen.

Um dennoch abzusichern, dass dieses Messmodell mit zwei getrennten latenten Faktoren für FSK und IM einem Ein-Faktor-Modell zu bevorzugen war, wurden beide Varianten geschätzt und hinsichtlich des Modellfits verglichen (Abbildung 17). Wie zu erwarten, erwies sich das Ein-Faktoren-Modell in beiden Domänen als weniger passend als das Zwei-Faktoren-Modell (Deutsch: CMIN = 69.92, df = 6, p = .00; GFI = .91; CFI = .84; RMSEA = .20 (.16 - .24); AIC = 99.92; Mathematik: CMIN = 102.85, df = 6, p = .00; GFI = .88; CFI = .87; RMSEA = .24 (.20 - .29); AIC = 132.85). Somit wurden FSK und IM in den folgenden Analysen als zwei getrennte Faktoren modelliert.

Abbildung 17: Einfaktorielles Messmodell der non-kognitiven Faktoren



Anmerkungen: FSK1-3: Items zur Erfassung des Fähigkeitsselbstkonzepts; IM1-3: Items zur Erfassung der intrinsischen Motivation; MOT: latenter Motivationsfaktor; e1-6: Fehlerterme.

Im Anschluss an die Überprüfung des Messmodells der non-kognitiven Faktoren erfolgte die Modellierung der Prognosemodelle. Diese werden im folgenden Kapitel dargestellt.

4.3.2 Prognosemodell der kognitiven und non-kognitiven Faktoren und des SES als Prädiktoren des Schulerfolgs

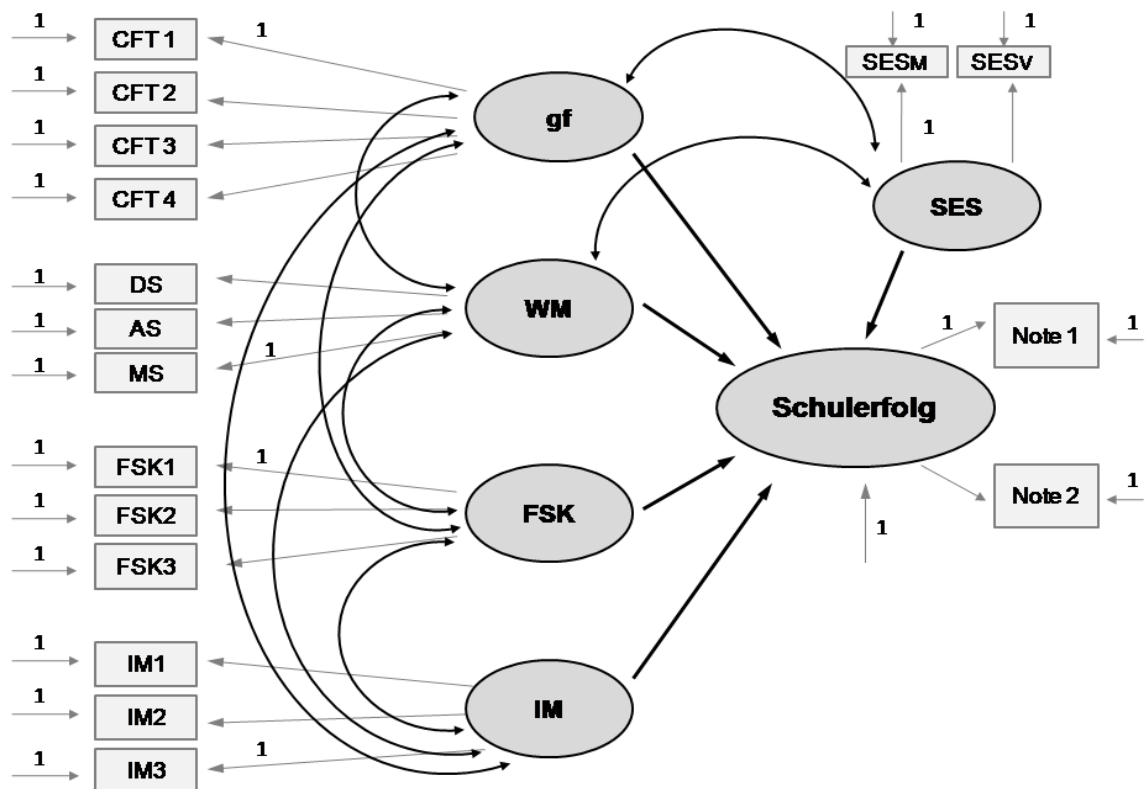
Zur Untersuchung der relativen Bedeutung kognitiver und non-kognitiver Prädiktoren bei der Vorhersage der Schulnoten in Deutsch und Mathematik wurde ein Vorhersagemodell geschätzt, in welchem gleichzeitig WM und gf als kognitive Prädiktoren und FSK und IM als non-kognitive Einflussfaktoren eingingen. Zusätzlich wurde der höchste Bildungsabschluss der Eltern als Indikator des sozioökonomischen Status der Familie modelliert. Zwischen allen kognitiven und non-kognitiven Prädiktoren wurde eine Korrelation zugelassen. Es wurden außerdem Korrelationen zwischen SES und den kognitiven Faktoren zugelassen (Abbildung 18).

Es wurden zwei separate Modelle für Deutsch und Mathematik geschätzt. So gingen in die Schätzung für Deutsch und Mathematik die entsprechenden Noten, sowie die fachspezifischen Motivationsitems ein.

Zunächst wurde die Normalverteilungsannahme überprüft. Der Mardia-Test indizierte mit einem Wert für die multivariate Kurtosis von 11.82 in Deutsch bzw. 24.54 in Mathematik und einem c.r.-Wert von 3.88 bzw. 8.05 eine Verletzung der multivariaten Normalverteilungsannahme in beiden Fächern. Dennoch lagen Schiefe und Kurtosis der einzelnen Items bzw. Subskalen innerhalb der von West, Finch und Curran (1995) angegebenen Grenzen von Schiefe $< |2|$ und Kurtosis $< |7|$, so dass gemäß der von Bühner (2006) empfohlenen Vorgehensweise weiterhin die ML-Methode Anwendung fand, welche sich bei einer Stichprobengröße von $N > 100$ als „ausgesprochen robust gegenüber Verletzungen der Normalverteilungsannahme“ (S. 251) erwiesen hat. Da dennoch im Zuge der ML-Schätzung von einem erhöhten χ^2 – Wert auszugehen war, wurde jeweils eine p-Wert Korrektur mittels der Bollen-Stine-Bootstrap Prozedur durchgeführt (Byrne, 2001), welche im Rahmen der Darstellung des Modellfits berichtet wird.

Die Ergebnisse dieser Schätzungen werden im Folgenden getrennt nach Fächern dargestellt.

Abbildung 18: Prognosemodell der kognitiven und non-kognitiven Faktoren und des SES



Anmerkungen: Aus Gründen der Übersichtlichkeit wurden die Fehlerterme nur angedeutet.

4.3.2.1 Prognosemodell in Deutsch

Das Prognosemodell im Fach Deutsch (Modell 5) ergab einen guten Gesamtfit (CMIN: 124.36, df = 103, p = .07; Bollen-Stine p = .15; GFI = .95; CFI = .98; RMSEA = .03 (.0 - .04); AIC = 224.36). Es erwiesen sich jedoch nicht die kognitiven Prädiktoren als einflussreichste Faktoren. Denn mit einem standardisierten Pfadgewicht von $\beta_{FSK} = .84$ stellte das fachspezifische FSK den stärksten Prädiktor dar. Intelligenz erklärte ebenfalls einen deutlichen Anteil der Kriteriumsvarianz, mit einem standardisierten Pfadgewicht von $\beta_{gf} = .37$, welches ebenfalls hoch signifikant war. Die Größe des Konfidenzintervalls deutete jedoch auf einen großen Standardfehler dieser Schätzung hin, was die Stabilität dieses Wertes in Frage stellte. Sowohl WM als auch IM leisteten keinen signifikanten Beitrag zur Vorhersage ($\beta_{WM} = -.18$, $\beta_{IM} = -.03$). Hiermit veränderte sich die Bedeutung des WM-Faktors durch Hinzunahme der non-kognitiven Prädiktoren, da in den Modellen, die ausschließlich kognitive Prädiktoren

einschlossen, WM einen signifikanten spezifischen Varianzanteil im Kriterium aufklärte, während gf unterhalb der Signifikanzgrenze lag.

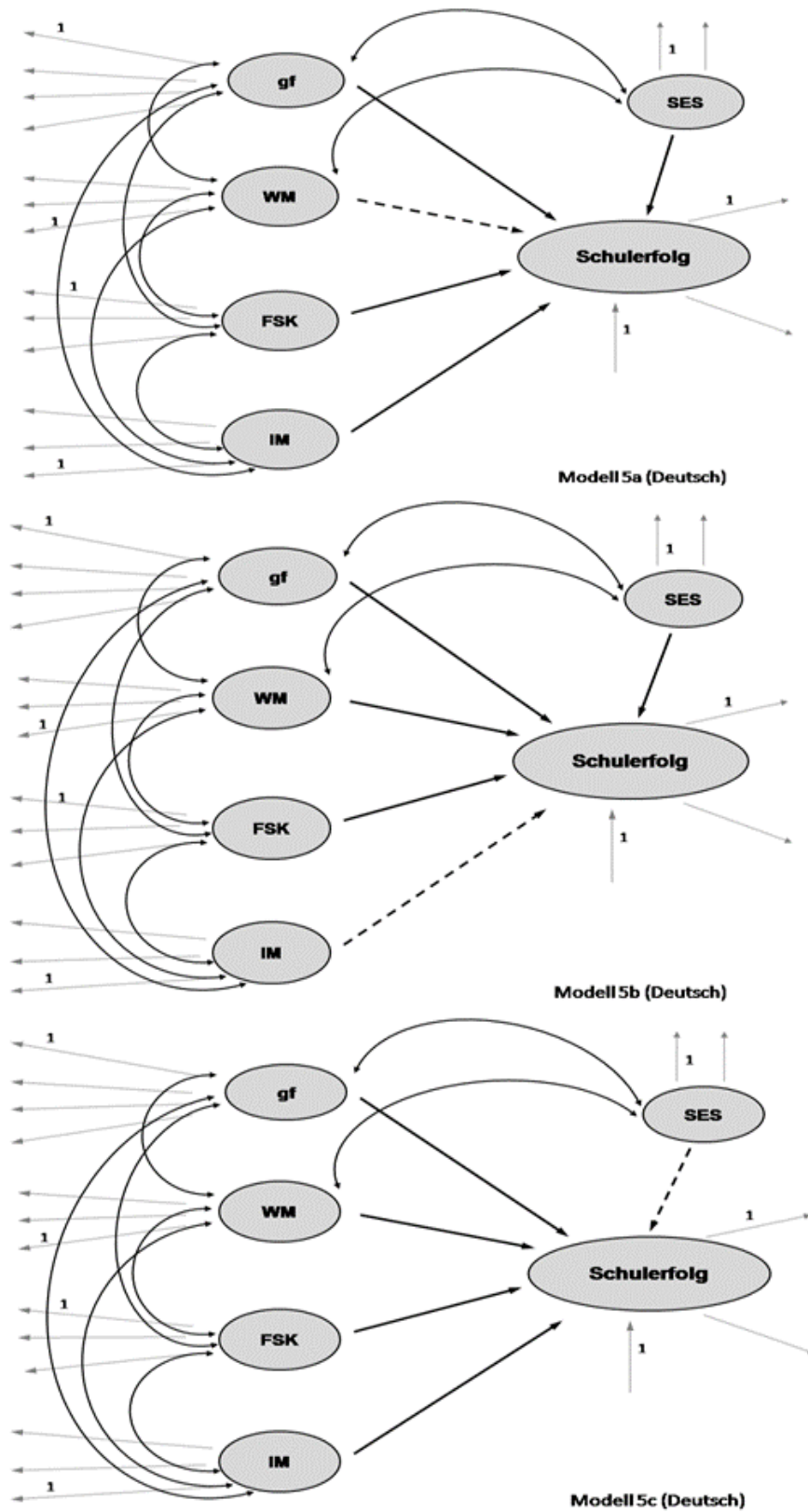
Der elterliche Bildungsabschluss klärte ebenfalls einen spezifischen Varianzanteil im Kriterium auf ($\beta_{\text{SES}} = .19$), wobei das 95%-ige Bootstrap-Konfidenzintervall durch Einschluss des Wertes null darauf hindeutete, dass diese Schätzung nicht stabil oberhalb der Signifikanzgrenze lag.

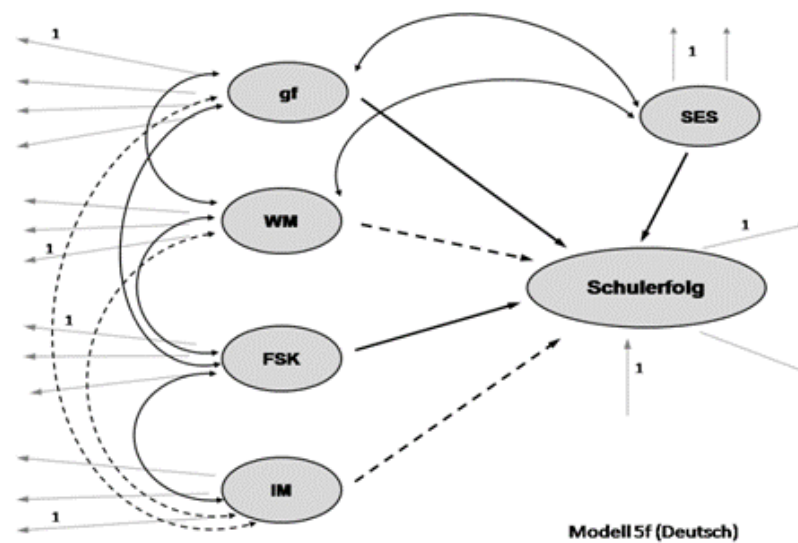
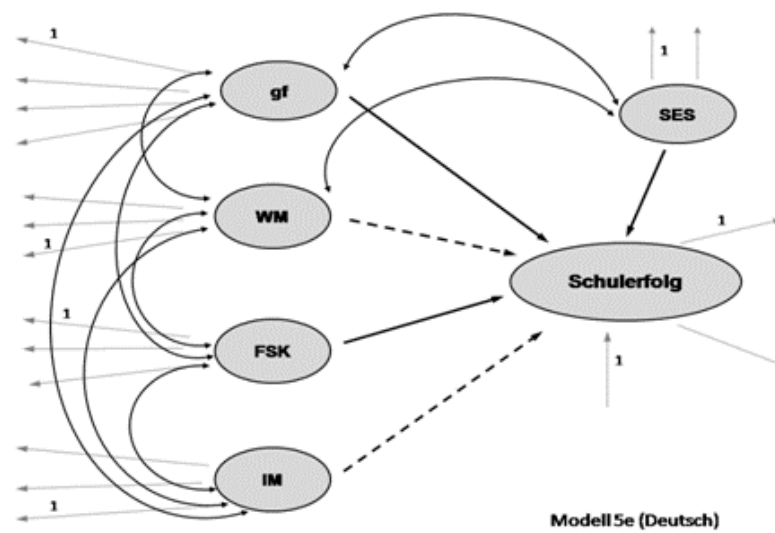
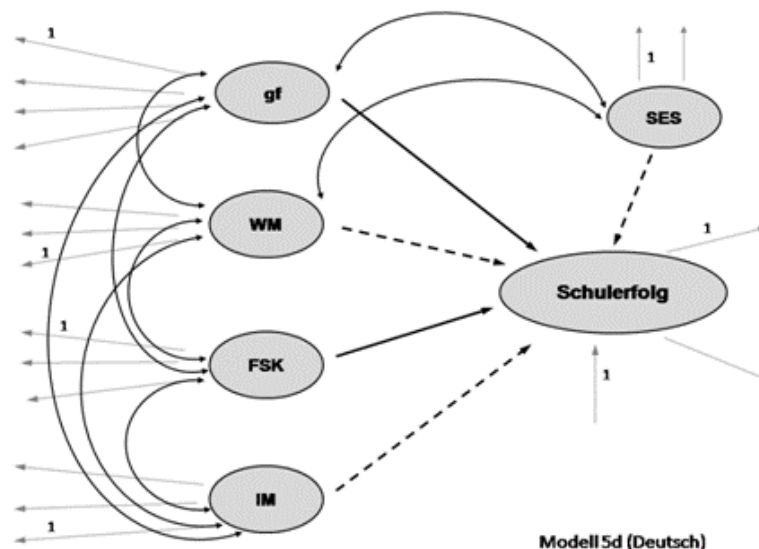
Die Betrachtung der Korrelationen zwischen den Prädiktoren ließ das Muster erkennen, dass FSK und IM mit $r = .56$ einen deutlichen Zusammenhang aufwiesen, wobei FSK auch mit WM assoziiert war, nicht jedoch mit gf. So ergab sich eine Korrelationen von $r = .46$ zwischen FSK und WM. IM hingegen wies keine signifikanten Assoziationen zu den kognitiven Prädiktoren auf. WM und gf korrelierten zu $r = .72$.

Der elterliche Bildungsabschluss wies Korrelationen von $r = .40$ zu gf und von $r = .22$ zu WM auf. Damit spiegelten die latenten Korrelationen das Muster der manifesten Zusammenhänge wider und sprachen für eine stärkere Beziehung des SES zu gf als zu WM. Insgesamt wurden 77% der Varianz im Kriterium aufgeklärt.

In einem nächsten Schritt wurden sukzessive nicht signifikante Pfade und Kovarianzen gleich null gesetzt, um genestete Modelle gegen das vollständige Modell zu testen, so dass überprüft werden konnte, ob ein parsimonischeres Modell dem vollständigen Modell vorzuziehen wäre. So wurde zunächst der Pfad von WM auf Schulleistung fixiert (Modell 5a), in einem weiteren Modell der Pfad von IM auf das Kriterium (Modell 5b). In Modell 5c wurde der Pfad von SES zu Schulerfolg auf null gesetzt. Modell 5c fittete zwar gemäß des χ^2 -Differenztests signifikant schlechter als das vollständige Modell. Da jedoch kein Fit-Index mit Ausnahme des χ^2 -Wertes eine Veränderung erfuhr und selbst der χ^2 -Wert mit einem $p = .05$ genau auf der Signifikanzgrenze lag, wurde ein Modell überprüft, in welchem sowohl WM und IM als auch SES auf null fixiert wurden (5d). Allerdings fittete auch dieses Modell signifikant schlechter. Da die Modelle 5a und 5b im Gegensatz zu Modell 5c und 5d in einem nicht signifikant schlechteren Fit mündeten, wurde anschließend ein weiteres Modell geschätzt, in welchem gleichzeitig die Pfade von WM und IM auf null gesetzt wurden (Modell 5e). Schließlich erfolgte die Testung eines Modells mit auf null gesetzten Pfaden von WM und IM auf die Schulnote, sowie mit fixierten Kovarianzen zwischen IM und den beiden kognitiven Prädiktoren (Modell 5f). Die schrittweise Fixierung der Pfade wird in Abbildung 19 illustriert.

Abbildung 19: Schematische Darstellung der schrittweisen Fixierung nicht signifikanter Pfade in Deutsch

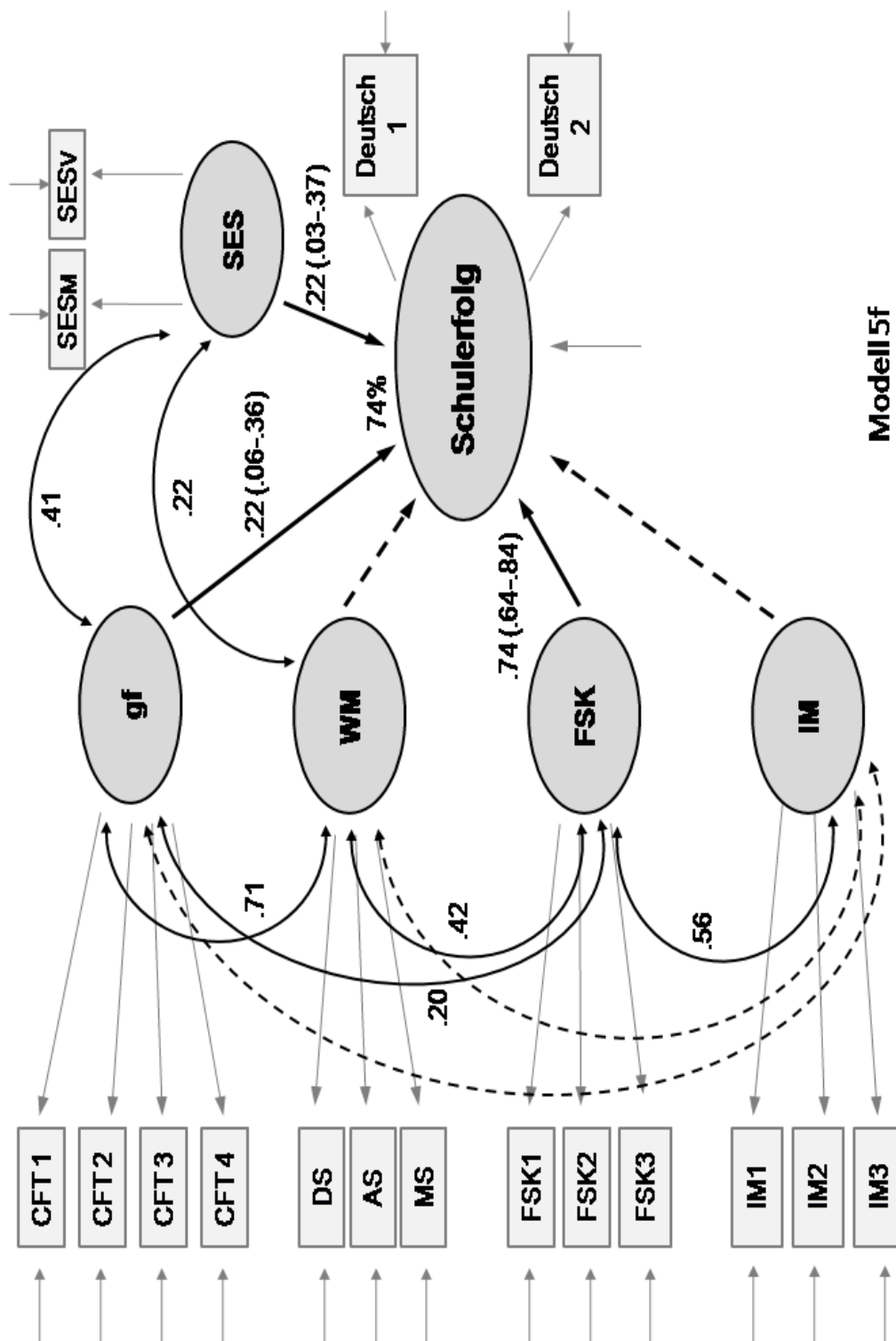




Anmerkungen: durchgezogene Linien: signifikante Pfade und Korrelationen; gestrichelte Linien: nicht signifikante Pfade und Korrelationen; aus Gründen der Übersichtlichkeit wurden die Fehlerterme nur angedeutet.

Alle auf diese Weise genesteten Modelle zeigten einen nicht signifikant schlechteren Modellfit (Tabelle 19), so dass schließlich dem parsimonischsten Modell (Modell 5f) der Vorzug zu geben war (CMIN: 4.20, df =4, p=.38). Dieses Modell ist in Abbildung 18 dargestellt. Während FSK mit einem Pfadgewicht von $\beta_{FSK} = .74$ (.64 - .84) stärkster Prädiktor blieb, ergaben sich sowohl für gf als auch für SES Werte von $\beta = .22$ (gf: .06 - .36; SES: .03 - .37). Die deutlichste Korrelation fand sich zwischen WM und gf mit einem Wert von $r = .71$, gefolgt von der Beziehung zwischen IM und FSK, zwischen denen ein Wert von $r = .56$ erreicht wurde. Auch WM hing sehr deutlich mit FSK zusammen ($r = .42$), während sich die Beziehung zwischen FSK und gf zwar als schwächer darstellte ($r = .20$), nun allerdings oberhalb der Signifikanzgrenze lag. Dieses Muster kehrte sich im Hinblick auf die Zusammenhänge zum SES um, denn hier fand sich ein höherer Wert für gf ($r = .41$) im Vergleich zu WM ($r = .22$). Insgesamt erklärte dieses Modell 74% der Kriteriumsvarianz.

Abbildung 18: Parsimonischstes Prognosemodell der kognitiven und non-kognitiven Faktoren, sowie des SES in Deutsch (Modell 5f)



Anmerkungen: N = 278; durchgezogene Linien: signifikante Pfade und Korrelationen; gestrichelte Linien: nicht signifikante Pfade und Korrelationen; Werte in Klammern beziehen sich auf das 95%-Bootstrap-Konfidenzintervall. Aus Gründen der Übersichtlichkeit wurden die Fehlerterme nur angedeutet.

4.3.2.2 *Prognosemodell in Mathematik*

Auch das Prognosemodell für Mathematik mit zwei kognitiven, zwei non-kognitiven Prädiktoren und elterlichem SES (Modell 6) ergab einen hervorragenden Gesamtfit (CMIN: 110.42, df = 103, p = .29; GFI = .96; CFI = .99; RMSEA = .02 (.0 - .04); AIC= 210.42). Der Bollen-Stine korrigierte p-Wert belief sich auf p = .41 und bestätigte damit die Gültigkeit des Modells.

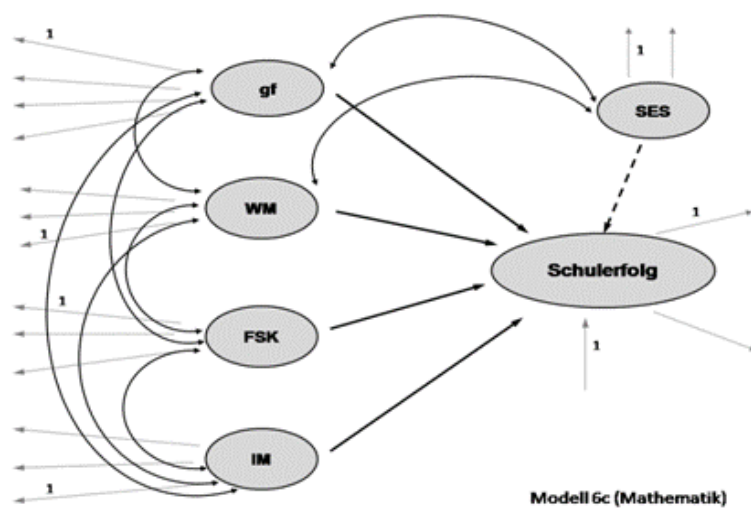
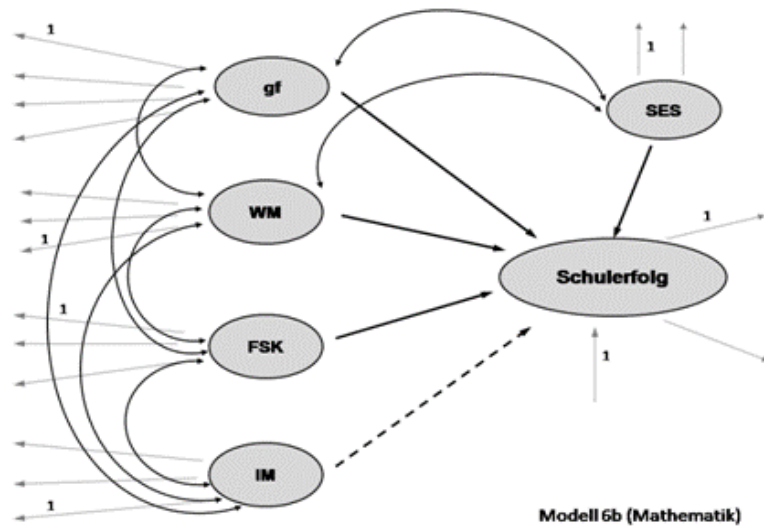
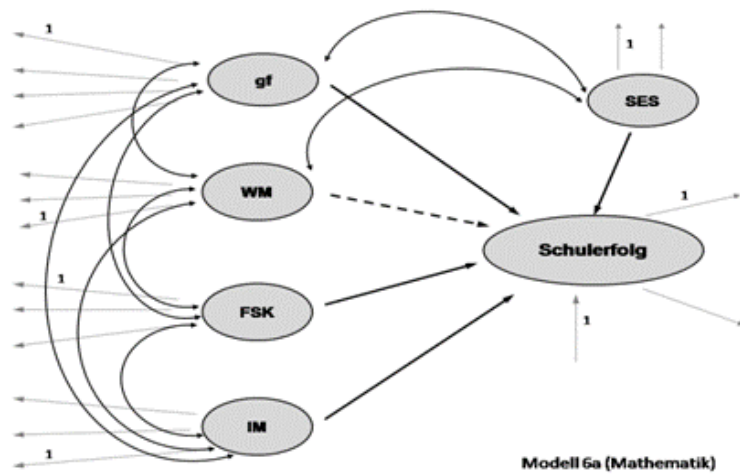
Hinsichtlich der relativen Bedeutung der einzelnen Prädiktoren bestätigte sich die prädiktive Kraft des fachspezifischen FSKs, welches mit einem standardisierten Pfadgewicht von $\beta_{FSK} = .62$ den stärksten Wert erreichte. Für gf ergab sich ein standardisiertes Pfadgewicht von $\beta_{gf} = .45$. Die Gewichte von IM ($\beta_{IM} = -.12$) und WM ($\beta_{WM} = .05$) lagen unterhalb der Signifikanzgrenze. Auch der Einfluss des SES erwies sich in Mathematik mit einem standardisierten Pfadgewicht von $\beta_{SES} = .11$ als nicht signifikant.

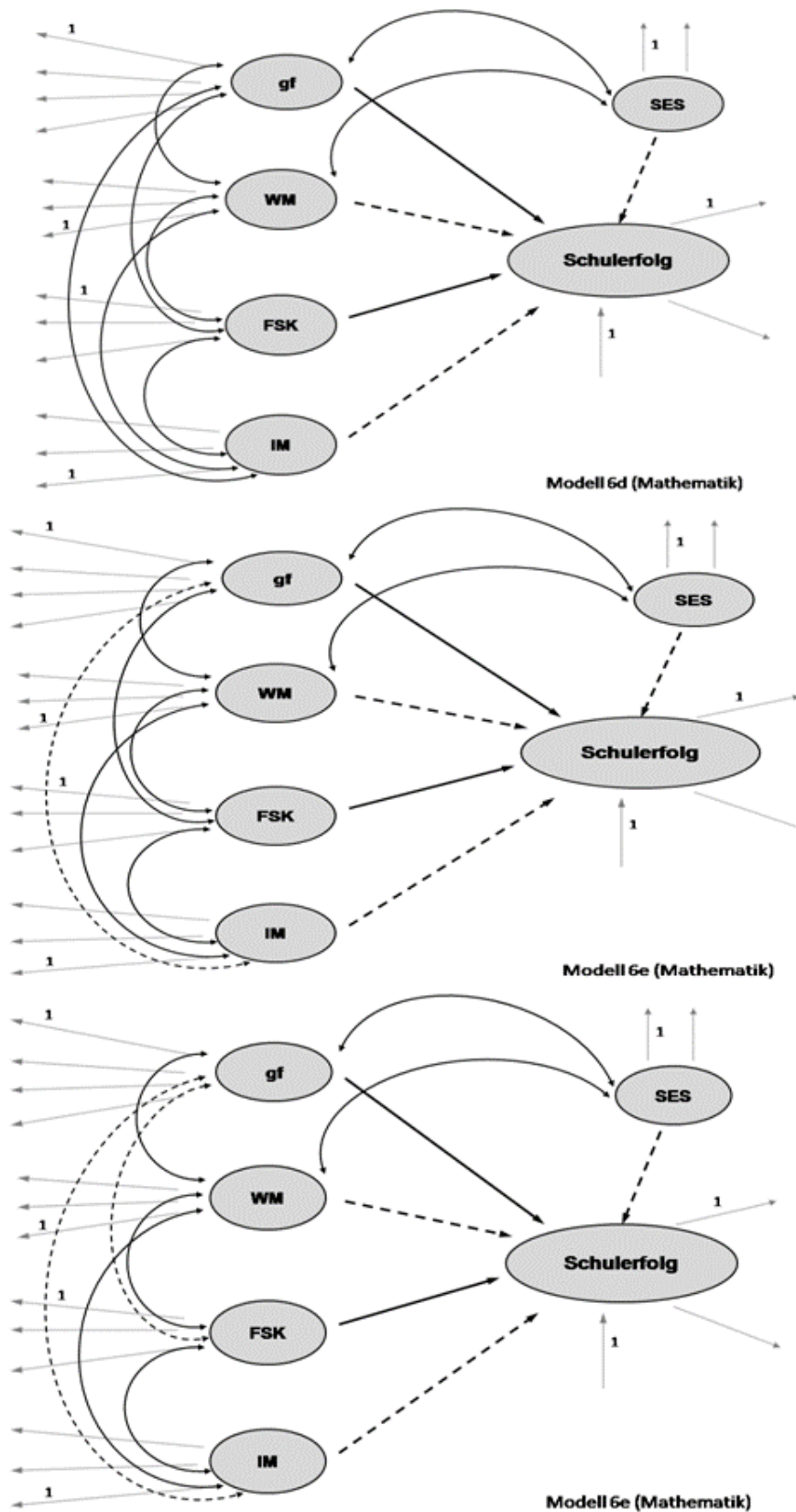
Im Hinblick auf die korrelativen Zusammenhänge zwischen den Prädiktoren, fand sich eine Korrelation von $r = .73$ zwischen gf und WM, eine Korrelation von $r = .41$ zwischen FSK und WM sowie ein Zusammenhang zwischen FSK und IM von $r = .73$. Wie auch im Fach Deutsch korrelierte FSK in Mathematik nicht mit gf. Allerdings ergab sich eine signifikante Korrelation zwischen WM und IM von $r = .21$, ein konsistenter Befund mit Bezug auf die manifesten Korrelationen.

SES hing mit $r = .43$ deutlicher mit gf zusammen als mit WM ($r = .23$). Insgesamt konnten 66% der Varianz im Kriterium durch Modell 6 aufgeklärt werden.

Im Anschluss an die Schätzung des vollständigen Modells 6 wurden sukzessive nicht signifikante Pfade bzw. Kovarianzen gleich null gesetzt, um parsimonischere Modelle schrittweise gegen das vollständige Modell testen zu können (siehe Abbildung 20). In Modell 6a wurde zunächst lediglich das Pfadgewicht von WM auf das Kriterium gleich null gesetzt. Analog hierzu wurden in Modell 6b der Pfad von IM auf das Kriterium fixiert, bzw. in Modell 6c der Pfad von SES zum Kriterium. Modell 6d umfasste nun drei Restriktionen, nämlich die Fixierung der Pfade von WM, IM und SES auf das Kriterium. Schließlich wurden in Modell 6e zusätzlich die Kovarianz zwischen IM und gf auf null fest gesetzt und in Modell 6f auch die Kovarianz zwischen FSK und gf. Das parsimonischste Modell 6f (Abbildung 21) zeigte einen nicht signifikant schlechteren Fit im Vergleich zum Ausgangsmodell (CMIN: 7.50, p = .18), so dass dieses Modell dem vollständigen Modell vorzuziehen war.

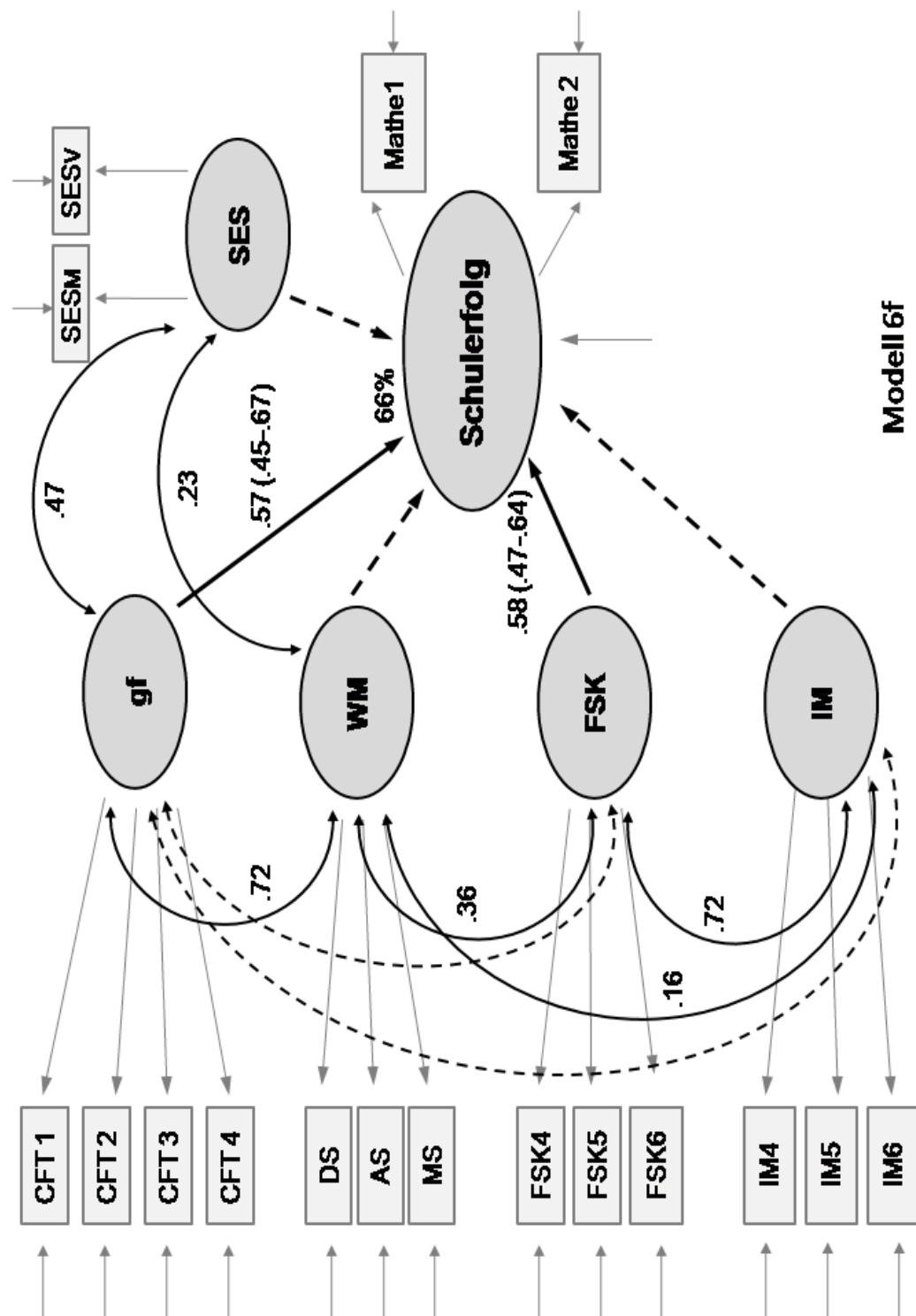
Abbildung 20: Schematische Darstellung der schrittweisen Fixierung nicht signifikanter Pfade in Mathematik





Anmerkungen: durchgezogene Linien: signifikante Pfade und Korrelationen; gestrichelte Linien: nicht signifikante Pfade und Korrelationen; aus Gründen der Übersichtlichkeit wurden die Fehlerterme nur angedeutet.

Abbildung 21: Parsimonischstes Prognosemodell der kognitiven und non-kognitiven Faktoren, sowie des SES in Mathematik (Modell 6f)



Anmerkungen: N = 278; durchgezogene Linien: signifikante Pfade und Korrelationen; gestrichelte Linien: nicht signifikante Pfade und Korrelationen; Werte in Klammern beziehen sich auf das 95%-Bootstrap-Konfidenzintervall; aus Gründen der Übersichtlichkeit sind die Fehlerterme nur angedeutet.

Es trugen hier also lediglich gf und FSK zur Vorhersage der Mathematiknote bei. Diese erwiesen sich nun als gleich starke Prädiktoren mit standardisierten Pfadgewichten von $\beta_{gf} = .57$ (.45 - .67) und $\beta_{FSK} = .58$ (.47 - .64). Im Gegensatz zum Fach Deutsch konnte FSK nun also keine dominierende Stellung mehr einnehmen. Der Einfluss des elterlichen SES verringerte sich ebenfalls, da dieser Faktor ohne Verschlechterung der Schätzung aus der Modellierung entfernt werden konnte. Zwischen SES und gf ergab sich eine Korrelation von $r = .47$ gefunden, WM korrelierte zu $r = .23$ mit SES. Zwischen gf und WM lag eine Korrelation von $r = .72$ vor. WM korrelierte zu $r = .36$ mit FSK und mit $r = .16$ mit IM. Zwischen FSK und IM lag eine Korrelation von $r = .72$ vor. Insgesamt wurden 66% der Kriteriumsvarianz durch Modell 6f aufgeklärt.

Die folgende Tabelle 19 fasst die Fit-Indizes der Modelle der kognitiven und non-kognitiven Prädiktoren sowie des SES zusammen.

Tabelle 19: Fit-Indizes der Prognosemodelle mit kognitiven und non-kognitiven Faktoren, sowie SES für Deutsch und Mathematik

	CMIN	df	p	GFI	CFI	RMSEA	AIC	CMIN*(df,p)
Deutsch								
Modell 5	124.36	103	.07	.95	.98	.03 (.00 - .04)	224.36	
Modell 5a	125.42	104	.07	.95	.98	.03 (.00 - .04)	223.42	1.06 (1,.30)
Modell 5b	124.44	104	.08	.95	.98	.03 (.00 - .04)	222.44	.08 (1,.78)
Modell 5c	128.48	104	.05	.95	.98	.03 (.00 - .04)	226.48	4.12 (1,.04)
Modell 5d	134.75	106	.03	.95	.98	.03 (.01 - .05)	228.75	10.40 (3,.02)
Modell 5e	125.42	105	.08	.95	.98	.03 (.00 - .04)	221.42	1.06 (2,.59)
Modell 5f	128.56	107	.08	.95	.98	.03 (.00 - .04)	220.56	4.20 (4,.38)
Mathematik								
Modell 6	110.42	103	.29	.96	.99	.02 (.00- .04)	210.42	
Modell 6a	110.49	104	.31	.96	1	.01 (.00 - .03)	208.49	.08 (1,.78)
Modell 6b	111.67	104	.29	.96	.99	.02 (.00- .04)	209.67	1.25 (1,.26)
Modell 6c	112.45	104	.27	.96	.99	.02 (.00- .04)	210.45	2.04 (1,.15)
Modell 6d	115.85	106	.24	.95	.99	.02 (.00- .04)	209.85	5.43 (3,.15)
Modell 6e	117.34	107	.23	.95	.99	.02 (.00- .04)	209.34	6.92 (4,.14)
Modell 6f	117.92	108	.24	.95	.99	.02 (.00- .04)	207.92	7.51 (5,.18)

Anmerkung: N = 278; Modell 5/6: Vollständig; Modell 5/6a bis 5/6f: genested; CMIN*: Wert des χ^2 -Differenztests zum Vergleich genesteter Modelle.

Die in den bisherigen Modellen geschätzten Pfadgewichte sind als spezifische Varianzanteile des jeweiligen Prädiktors zu verstehen. Doch sowohl die Ergebnisse der Kommunalitätenanalysen in Kapitel 4.2.1, als auch die zahlreichen signifikanten Korrelationen wiesen auf große Anteile geteilter Varianz zwischen den Prädiktoren hin. Um nun spezifischer die Einflüsse der kognitiven und non-kognitiven Faktoren kontrastieren zu können und mehr über die relative Bedeutung der beiden Prädiktorgruppen zu erfahren, wurden im Anschluss Zwei-Ebenen-Modelle modelliert, getrennt für Deutsch und Mathematik (Abbildung 22). Die diesbezüglichen Befunde werden im folgenden Kapitel dargestellt.

4.3.3 Zwei-Ebenen-Modellierung der kognitiven und non-kognitiven Faktoren und des SES als Prädiktoren des Schulerfolgs

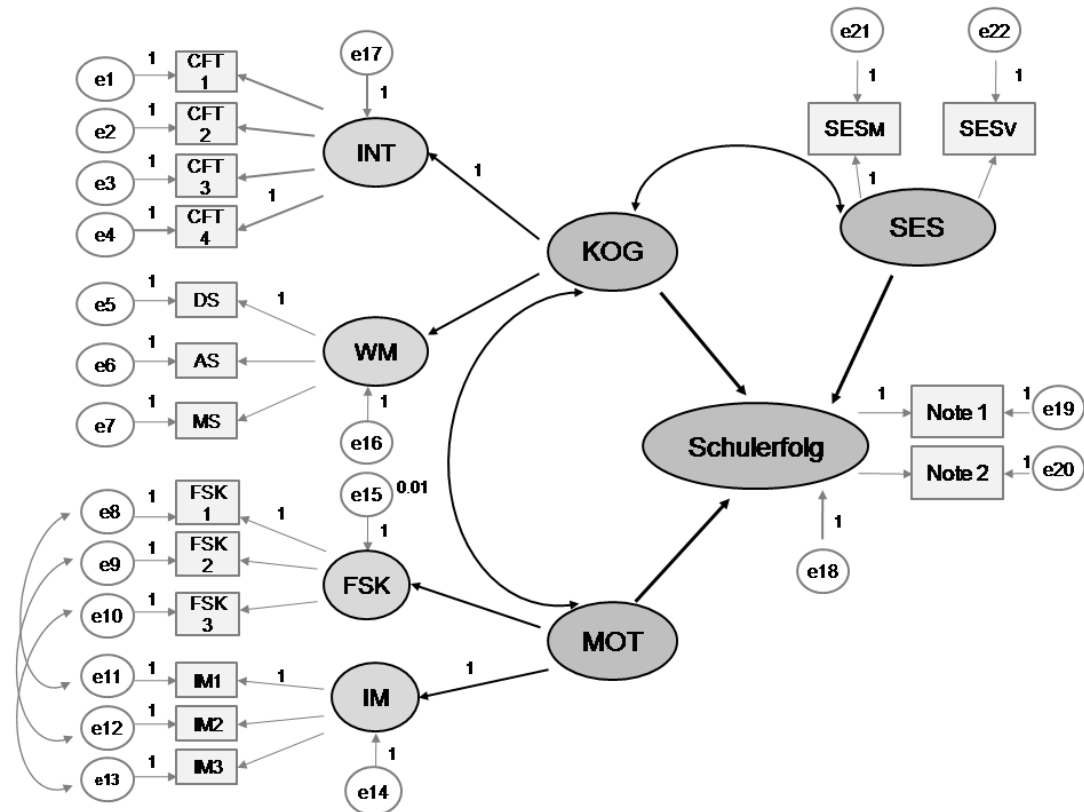
Analog zum Zwei-Ebenen-Modell der kognitiven Prädiktoren wurde in diesem Modell ein latenter Motivationsfaktor zweiter Ordnung modelliert, der auf einen latenten FSK- und einen latenten IM-Faktor erster Ordnung lud. Außerdem wurde ein latenter Kognitionsfaktor zweiter Ordnung geschätzt, welcher auf einen latenten WM- und einen latenten gf-Faktor erster Ordnung lud (Abbildung 22). Die beiden Faktoren zweiter Ordnung wurden nun zur Vorhersage der Schulnote herangezogen, wobei eine Korrelation zwischen dem Motivations- und dem Kognitionsfaktor zugelassen wurde. Die Tatsache, dass sowohl die kognitiven Variablen als auch die beiden non-kognitiven Variablen deutliche Korrelationen aufwiesen und damit gemeinsame Varianzanteile von substantieller Ausprägung anzunehmen waren, rechtfertigte diese Art der Modellierung.

Auch in diesem Modell wurde der sozioökonomische Status der Eltern durch Modellierung des höchsten elterlichen Bildungsabschlusses einbezogen und eine Korrelation zum Kognitionsfaktor zugelassen.

Es ist zu beachten, dass eine negative Fehlervarianz im latenten FSK-Faktor auftrat, wodurch einflussreiche Extremwerte oder eine Fehlspezifikationen indiziert werden können. Die Tatsache, dass jedoch vor der Modelltestung multivariate Ausreißer bereits entfernt wurden, ließ den Schluss zu, dass der Einfluss von Ausreißern als Erklärung eher unwahrscheinlich war. Da außerdem der Wert der negativen Fehlervarianz nur minimal und nicht signifikant war (-.04) und das Modell trotz dieses Umstands konvergierte, wurde die Fehlervarianz gemäß der von Chen, Bollen,

Paxton, Curran & Kirby (2001) empfohlenen Vorgehensweise in beiden Fächern auf den Wert von 0,01 fixiert.

Abbildung 22: Zwei-Ebenen-Modell der kognitiven und non-kognitiven Faktoren und des SES



Anmerkungen: die Fehlervarianz des latenten FSK-Faktors wurde auf 0,01 fixiert.

4.3.3.1 Zwei-Ebenen-Modell Deutsch

Die Modellschätzung für Deutsch (Modell 7) bestätigte die Stärke der motivationalen Konstrukte im Hinblick auf die Vorhersage der Deutschnote im Vergleich zu den kognitiven Prädiktoren. So lud der latente Kognitionsfaktor mit einem standardisierten Gewicht von $\beta_{\text{KOG}} = .23$ (.01 - .41) zwar hoch signifikant auf das Kriterium, doch der latente Motivationsfaktor überstieg die Vorhersagekraft des Kognitionsfaktors mit einem Wert von $\beta_{\text{MOT}} = .72$ (.56 - .84). Die Vertrauensintervalle ergaben keine Überlappung der standardisierten Pfadgewichte. Der Kognitionsfaktor lud mit $\beta = .84$ (.63 - 1) auf den latenten WM-Faktor und mit $\beta = .87$ (.61 - 1) auf den latenten gf-

Faktor. Der Motivationsfaktor zweiter Ordnung lud mit $\beta = .98$ (.97 - .99) auf den latenten FSK-Faktor und mit $\beta = .56$ (.34 - .69) auf IM. Damit lag eine stärkere Assoziation des Faktors MOT zu FSK als zu IM vor. Zwischen den Faktoren zweiter Ordnung konnte eine Korrelation von $r = .31$ festgestellt werden. SES korrelierte zu $r = .43$ mit dem Kognitionsfaktor. Der SES erreicht ebenfalls ein signifikantes Pfadgewicht von $\beta_{SES} = .23$ (.02 - .37), wobei sich das Konfidenzintervall nahezu vollständig mit dem Intervall des KOG-Faktors deckte. Insgesamt wurden mit Hilfe des Modells 7 77% der Kriteriumsvarianz aufgeklärt. Der Modellfit erwies sich als schlechter verglichen mit den vorherigen Modellen, insgesamt war er jedoch als akzeptabel zu bewerten (CMIN: 147.19, 109, .00; Bollen-Stine $p = .03$; GFI = .94; CFI = .97; RMSEA = .04 (.02 - .05), AIC = 237.19).

Ein genestetes Modell, in welchem der Pfad von SES zu Schulerfolg auf null gesetzt wurde, lieferte einen signifikant schlechteren Fit, weshalb das vollständige Modell zu bevorzugen war (CMIN: 9.61; $p = .00$) (siehe Abbildung 23).

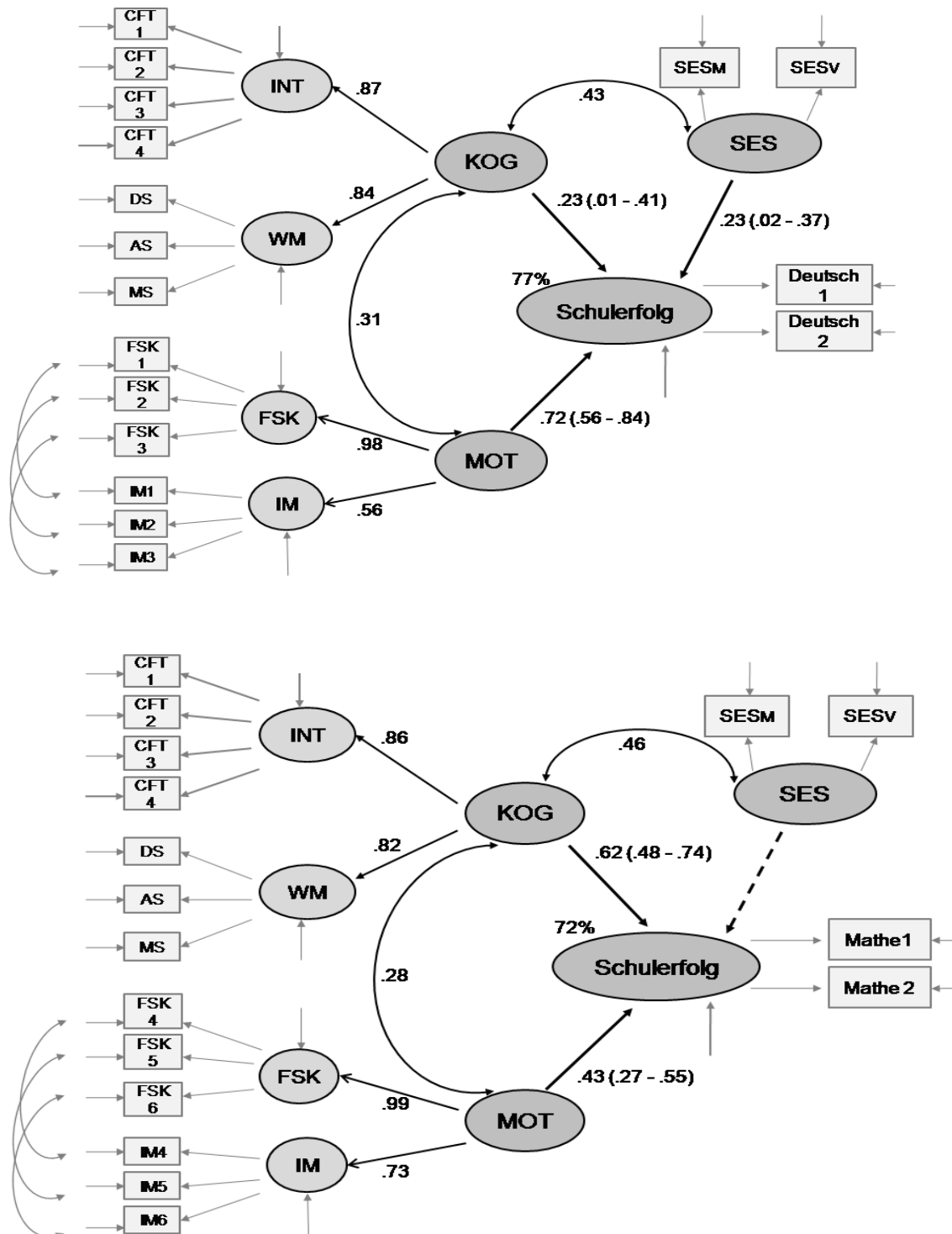
Tabelle 20 fasst die Fit-Indizes der Zwei-Ebenen-Modelle in Deutsch und Mathematik zusammen.

Tabelle 20: Fit-Indizes der Zwei-Ebenen-Modelle mit kognitiven und non-kognitiven Faktoren, sowie SES in Deutsch und Mathematik

	CMIN	df	p	GFI	CFI	RMSEA	AIC	CMIN*(df,p)
Deutsch								
Modell 7	147.19	108	.01	.94	.97	.04 (.02 - .05)	237.19	
Modell 7a	156.81	109	.00	.94	.96	.04 (.02 - .05)	244.81	9.61 (1,.00)
Mathematik								
Modell 8	129.16	108	.08	.95	.99	.03 (.00 - .04)	219.16	
Modell 8a	131.60	109	.07	.95	.99	.03 (.00 - .04)	219.60	2.44 (1,.12)

Anmerkungen: N = 278; Modell 7/8: Vollständig; Modell 7/8a: SES constrained; CMIN*: Wert des χ^2 -Differenztests zum Vergleich genesteter Modelle.

Abbildung 23: Parsimonischstes Zwei-Ebenen-Modell der kognitiven und non-kognitiven Faktoren, sowie des SES in Deutsch und Mathematik (Modell 7/ Modell 8a)



Anmerkungen: N = 278; durchgezogene Linien: signifikante Pfade und Korrelationen; gestrichelte Linien: nicht signifikante Pfade und Korrelationen; Werte in Klammern beziehen sich auf das 95%-Bootstrap-Konfidenzintervall. Aus Gründen der Übersichtlichkeit wurden die Fehlerterme nur angedeutet.

Insgesamt bestätigte das Zwei-Faktoren-Modell das Ergebnismuster der vorherigen Modelle im Fach Deutsch dahingehend, dass nicht die kognitiven Prädiktoren als stärkste Prädiktoren in Erscheinung traten, sondern die motivationalen Faktoren, im besonderen FSK. Die prädiktiven Validitäten von SES und KOG erwiesen sich als vergleichbar.

4.3.3.2 Zwei-Ebenen-Modell Mathematik

Die Modellierung des Zwei-Ebenen-Modells in Mathematik (Modell 8) lieferte jedoch abweichende Befunde. Zunächst erwies sich der Modellfit für Mathematik insgesamt als besser verglichen mit Deutsch (CMIN 129.16, 108, .08; Bollen-Stine $p=.17$; GFI = .95; CFI=.99; RMSEA = .03 (.0-.04), AIC = 219.16). Außerdem ergab die Betrachtung der standardisierten Pfadgewichte, dass der latente Kognitionsfaktor sich nun als stärkerer Prädiktor erwies $\beta_{\text{KOG}} = .54$, verglichen mit dem Motivationsfaktor $\beta_{\text{MOT}} = .45$. Zwar konnte eine starke Überlappung der beiden Konfidenzintervalle festgestellt werden, dennoch lag das standardisierte Pfadgewicht des Kognitionsfaktors über dem des Motivationsfaktors.

SES trug mit $\beta_{\text{SES}} = .11$ nicht signifikant zur Vorhersage des Kriteriums bei. Zwischen SES und KOG ergab sich eine Korrelation von $r = .41$. MOT und KOG korrelierten zu $r = .28$. Der Kognitionsfaktor zweiter Ordnung lud in vergleichbarer Höhe auf WM und gf, nämlich zu $\beta = .84$ bzw. $\beta = .87$. Der latente Motivationsfaktor zweiter Ordnung lud mit $\beta = .99$ höher auf FSK als auf IM ($\beta = .73$). Insgesamt konnte 69% der Kriteriumsvarianz durch Modell 8 aufgeklärt werden.

Da sich SES in diesem Modell als nicht signifikanter Prädiktor erwies, wurde ein genestetes Modell getestet (Modell 8a), in welchem das Pfadgewicht von SES zum Kriterium gleich null gesetzt wurde. Dieses parsimonischere Modell fittete nicht signifikant schlechter (CMIN: 2.19, $p = .16$) als das vollständige Modell, so dass es diesem vorzuziehen war (Abbildung 23).

Die korrelativen Zusammenhänge erwiesen sich als stabil, wobei SES und KOG nun zu $r = .46$ korrelierten. Das standardisierte Pfadgewicht des Kognitionsfaktors erreichte einen Wert von $\beta_{\text{KOG}} = .62$ (.48 - .74). Das standardisierte Gewicht des Motivationsfaktors belief sich auf $\beta_{\text{MOT}} = .43$ (.27 - .55). Die Tatsache, dass die Konfidenzintervalle nur eine minimale Überlappung aufwiesen, untermauerte den Befund, dass sich im Fach Mathematik die Bedeutung der motivationalen Faktoren im direkten Vergleich zu den kognitiven Faktoren verringerte. Modell 8a erreichte eine Varianzaufklärung von 72%.

Zur Veranschaulichung der Ergebnisse werden die standardisierten Pfadgewichte und latenten Korrelationen der Modelle in Deutsch und Mathematik in Tabelle 21 zusammengefasst (Modelle 5 – 8). Die Faktorladungen der Indikatorvariablen sind Anhang A zu entnehmen.

Tabelle 21: Standardisierte Pfadgewichte und latente Korrelationen der Modelle mit kognitiven und non-kognitiven Faktoren, sowie SES getrennt nach Schulfach

			Deutsch		Mathematik	
			β	r	β	r
Vorhersagemodell						
gf	→	SE	.22 (.06 - .36)	-	.57 (.45 - .67)	-
WM	→	SE	0	-	0	-
SES	→	SE	.22 (.03 - .37)	-	0	-
FSK	→	SE	.74 (.64 - .84)	-	.58 (.47 - .64)	-
IM	→	SE	0	-	0	-
WM	↔	gf	-	.71	-	.72
WM	↔	SES	-	.22	-	.23
gf	↔	SES	-	.41	-	.47
FSK	↔	IM	-	.56	-	.72
FSK	↔	WM	-	.42	-	.36
FSK	↔	gf	-	.20	-	0
IM	↔	WM	-	0	-	.16
IM	↔	gf	-	0	-	0
Zwei-Ebenen-Modell						
KOG	→	SE	.23 (.01 - .41)	-	.62 (.48 - .74)	-
MOT	→	SE	.72 (.56 - .84)	-	.43 (.27 - .55)	-
SES	→	SE	.21 (.02 - .37)	-	0	-
KOG	→	WM	.84 (.63 - 1)	-	.82 (.62 - .95)	-
KOG	→	gf	.87 (.61 - 1)	-	.86 (.71 - .97)	-
MOT	→	FSK	.98 (.97 - .99)	-	.99 (.98 - .99)	-
MOT	→	IM	.56 (.34 - .69)	-	.73 (.59 - .80)	-
KOG	↔	MOT	-	.31	-	.28
KOG	↔	SES	-	.43	-	.46

Anmerkungen: β : standardisiertes Pfadgewicht; r: latente Korrelation; gf: latenter gf-Faktor; WM: latenter WM-Faktor; SES: latenter SES-Faktor; FSK: latenter FSK-Faktor; IM: latenter IM-Faktor; KOG: latenter Kognitionsfaktor; MOT: latenter Motivationsfaktor; SE: Schulerfolg; Werte beziehen sich auf das jeweils parsimonischste Modell, getrennt nach Schulfach.

4.3.4 Zusammenfassung

Die Zusammenschau der vorgestellten Modellierungen bezüglich der relativen Bedeutung kognitiver und non-kognitiver Variablen und des SES, ergab für das Fach Deutsch folgendes Bild. Konsistent über alle Modelle hinweg erwiesen sich die motivationalen Faktoren als stärkere Prädiktoren verglichen mit WM, gf und SES. Besonders das Fähigkeitsselbstkonzept nahm eine dominierende Stellung ein. Intrinsische Motivation konnte keine spezifischen Varianzanteile aufklären. Es ist jedoch zu beachten, dass FSK und IM einen substantiellen Anteil an Varianz teilten, was aus den hohen korrelativen Zusammenhängen zu schließen ist.

Im Hinblick auf die kognitiven Prädiktoren zeigte sich, dass lediglich gf einen spezifischen Beitrag zur Varianzaufklärung leisten konnte, während sich das Pfadgewicht von WM zur Schulnote als nicht signifikant erwies. Durch Bündelung der gemeinsamen Varianz im Zwei-Ebenen-Modell trugen jedoch beide Faktoren zur Vorhersage bei.

Der elterliche Bildungsabschluss leistete ebenfalls einen signifikanten Beitrag zur Varianzaufklärung im Kriterium. Die Höhe des Pfadgewichts war vergleichbar mit dem Wert des Pfades von gf zum Kriterium, bzw. im Zwei-Ebenen-Modell mit dem Pfadgewicht des Kognitionsfaktors.

Hinsichtlich der korrelativen Beziehungen zeigte sich, dass sowohl WM und gf, als auch FSK und IM deutliche Assoziationen aufwiesen. Außerdem korrelierten die kognitiven Faktoren mit SES, wobei die Werte für WM niedriger ausfielen als für gf. Umgekehrt war FSK stärker mit WM assoziiert als mit gf. IM wies keine Korrelationen zu den kognitiven Prädiktoren auf.

Das Zwei-Ebenen-Modell zeigte zwar Einbußen im Modellfit, diese waren jedoch nur als mäßig zu bezeichnen, in Anbetracht der höheren Komplexität des Modells.

Im Fach Mathematik ergab die zusammenfassende Betrachtung der verschiedenen Modellierungen den konsistenten Befund, dass die non-kognitiven Prädiktoren, bzw. FSK, keine dominierende Rolle mehr einnehmen konnten. Im initialen Prognosemodell (Modell 6) erwiesen sich FSK und gf als gleich gewichtige Prädiktoren. Bündelte man jedoch die gemeinsame Varianz der kognitiven Variablen

bzw. der non-kognitiven Variablen in einem Zwei-Ebenen-Modell (Modell 8), überstieg die prädiktive Validität des Kognitionsfaktors die Vorhersagekraft des Motivationsfaktors.

Der elterliche Bildungsabschluss erwies sich im Unterschied zum Fach Deutsch als irrelevant für die Prädiktion der Mathematiknote.

Die Höhe der korrelativen Zusammenhänge zwischen den Prädiktoren änderte sich kaum im Vergleich zu Deutsch.

Der direkte Vergleich der Fit-Indizes zeigte, dass die komplexere Modellierung eines Zwei-Ebenen-Modells in Mathematik keine Einbußen bezüglich der Modellpassung nach sich zog. Somit konnte die Eignung dieses Modells zur direkten Kontrastierung der prädiktiven Validität kognitiver und non-kognitiver Faktoren als bestätigt betrachtet werden.

Im folgenden Kapitel erfolgt eine zusammenfassende Diskussion der Befunde.

V Diskussion

Die sich anschließende Diskussion dient der zusammenfassenden Betrachtung und Interpretation der dargestellten Ergebnisse. Dabei sollen die der Arbeit zugrunde liegenden Hypothesen bezüglich ihrer Gültigkeit bewertet werden.

Nach der Ergebnisdiskussion erfolgt eine kritische Betrachtung der vorliegenden Studie. Abschließend wird ein Ausblick auf zukünftige Forschungsfragen, sowie auf praktische Implikationen der Studie gegeben.

Bevor die Befunde der Strukturgleichungsanalysen zusammenfassend diskutiert werden, erfolgt vorab die Diskussion der Binnenstruktur der verschiedenen Prädiktorgruppen, welche für die Interpretation der weiteren Analysen, sowie die Bewertung der Hypothesen von Relevanz sind.

5.1 Erhebung und Binnenstruktur der kognitiven Fähigkeiten

5.1.1 Arbeitsgedächtnis

Während zur Messung der Intelligenz ein weit verbreitetes, standardisiertes Verfahren genutzt werden konnte (CFT 20-R), musste zur Messung der Arbeitsgedächtniskapazität auf Aufgaben zurückgegriffen werden, die speziell für die vorliegende Studie entwickelt bzw. aus bereits bestehenden Aufgaben adaptiert wurden. Die Analyse der Reliabilitäten ergibt für die Arbeitsgedächtnisaufgaben Werte im niedrigen bis moderaten Bereich ($r_{ttc} = .50 - .73$), wobei die AS-Aufgabe im Vergleich zur DS- und MS-Aufgabe den schlechtesten Wert aufweist ($r_{ttc} = .50$). Begründen lässt sich die mangelnde Reliabilität dieser Aufgabe zum Teil durch die Art des eingesetzten Stimulusmaterials, bzw. die Struktur der Aufgabe. So spielen in dieser Rechtschreibprozesse eine große Rolle, da die Kinder jeden Tiernamen ausschreiben mussten, was eine höhere Beanspruchung darstellt als beispielsweise das Schreiben von Ziffern, welches schneller vonstattengeht und einfacher ist, da keine orthographischen Regeln beachtet werden müssen. Damit stellen die DS- und

MS-Task reinere WM- Indikatoren dar, die durch weniger Störeinflüsse konfundiert sind, was sich in höheren Reliabilitäten widerspiegelt. Dass die DS-Aufgabe insgesamt für die Kinder einfacher zu bearbeiten ist als die AS- oder MS-Aufgabe, kann aus der Verteilung der Mittelwerte abgeleitet werden, da hier die DS-Aufgabe den höchsten Wert erreicht ($\bar{x} = 7.45$).

Weiterhin gilt bei der Betrachtung der Ergebnisse anzumerken, dass die Testung in Gruppen zu Einbußen in der Reliabilität führen kann. Möglicherweise hätte durch eine computergestützte Einzeltestung die Reliabilität gesteigert werden können, da eine solche Erhebungsmethode es erschwert, entgegen der Instruktion zu arbeiten, sich während der Darbietungszeit Notizen zu machen oder zu versuchen, die Angaben des Nachbarn abzulesen. Auch wenn durch eine möglichst kleine Gruppengröße und die Anwesenheit zweier Testleiter versucht wurde, Störeffekte zu kontrollieren, kann nicht ausgeschlossen werden, dass diese in Einzelfällen das Testergebnis beeinflusst haben.

Die drei WM-Aufgaben weisen nicht nur unterschiedlich hohe Reliabilitäten auf, sie unterscheiden sich außerdem hinsichtlich der Stärke der Interkorrelationen. Alle drei Aufgaben korrelieren zwar positiv und hoch signifikant, doch die höchste Korrelation liegt zwischen den Aufgaben DS und AS vor ($r = .39$). Dies erklärt sich durch die stärkere Fokussierung dieser beiden Aufgaben auf verbale Komponenten im Vergleich zur MS-Task, welche vornehmlich die nonverbale Domäne anspricht. Die Tatsache, dass DS und AS somit mehr Varianz untereinander als mit der MS-Task teilen, kann als Hinweis auf domänen-spezifische Mechanismen gewertet werden. Doch die insgesamt positiven Interkorrelationen verdeutlichen, dass Arbeitsgedächtnis zumindest zum Teil über die Domänen hinweg operiert. Einschränkend ist jedoch zu sagen, dass die hier eingesetzten Aufgaben keine strikte Trennung einzelner Domänen erlauben, da beispielsweise die AS-Aufgabe neben der verbalen Komponente auch visuell-räumlichen Charakter hat, da die Körpergröße der Tiere verarbeitet werden muss, um die Aufgabe richtig lösen zu können. Da jedoch nicht die Überprüfung eines spezifischen theoretischen Modells oder bestimmter Modellkomponenten Ziel der Untersuchung ist, ist dieser Aspekt für die Beantwortung der Fragestellung der vorliegenden Arbeit weniger problematisch.

Insgesamt betrachtet, zeigt sowohl die Tatsache, dass ein aus allen drei Aufgaben errechneter Summenscore eine Reliabilität von $r_{ttc} = .77$ aufweist, als auch der Nachweis eines starken gemeinsamen WM-Faktors (52.4% Varianzaufklärung), dass

die Bewertung der drei WM-Aufgaben als Indikatoren eines gemeinsamen Konstrukts stimmig ist. Auch die Faktorladungen des latenten Messmodells der kognitiven Fähigkeiten bestätigen den latenten WM-Faktor mit Ladungen zwischen $\lambda = .45$ und $.65$ (standardisierte Lösung).

Im nächsten Abschnitt werden die Befunde bezüglich der Binnenstruktur des erhobenen Intelligenzkonstrukts diskutiert.

5.1.2 Intelligenz

Wie bereits erwähnt wurde die Intelligenz mit Hilfe des CFT 20-R erfasst. Trotz des Einsatzes dieses standardisierten Verfahrens können auf Subtestniveau ebenfalls lediglich niedrige bis moderate Reliabilitäten festgestellt werden ($r_{ttc} = .47 - .67$). Möglicherweise werden die Ergebnisse auch hier durch die mit einer Gruppentestung verknüpften Störeffekte beeinflusst. Da jedoch im Manual des CFT 20-R keine Subtestreliabilitäten angegeben werden, die zu Vergleichszwecken hätten herangezogen werden können, wurde die Reliabilität der Gesamtskala (Summenscore über alle vier Subtests) mit dem im Manual berichteten Wert von $r = .92$ verglichen. Die Gesamtskala erreicht einen Wert von $r_{ttc} = .78$ und weicht somit vom Vergleichswert des CFT-Manuals ab. Dennoch ist dieser Wert als akzeptabel zu bezeichnen (Kline, 2005), besonders in Anbetracht der kleineren Stichprobengröße im Vergleich zu der dem CFT-Manual zugrunde liegenden Stichprobe ($N > 4000$). Da die weiterführenden Analysen entweder mit Faktorwerten durchgeführt wurden, bzw. die Subtests lediglich als Indikatoren eines latenten Faktors fungieren, ist die Reliabilität der Gesamtskala ausschlaggebender als die Subtestreliabilitäten.

Die unterschiedliche Reliabilität der vier Subtests spiegelt sich auch im Muster der Interkorrelationen wider. So variieren die Korrelationen zu Subtest 1, welcher mit einem Wert von $r_{ttc} = .47$ die niedrigste Reliabilität aufweist. Subtest 2, welcher ebenfalls einen eher niedrigen Reliabilitätswert erzielt ($r_{ttc} = .52$), korreliert daher niedriger mit Subtest 1 ($r = .24$) als Subtest 3 ($r = .40$), welcher die höchste Reliabilität aufweist ($r_{ttc} = .67$).

Zusammengefasst sprechen jedoch der Nachweis eines starken gf-Faktors (48.57% Varianzaufklärung), die positiven Interkorrelationen der Subtests, sowie die akzeptable Reliabilität der Gesamtskala für die Interpretation der vier Subtests als

Indikatoren eines gemeinsamen Intelligenzfaktors. Auch hier bekräftigen die im latenten Messmodell geschätzten standardisierten Faktorladungen diese Annahme, mit Werten zwischen $\lambda = .46$ und $.64$.

Im Anschluss an die zusammenfassende Betrachtung der Struktur der kognitiven Fähigkeiten erfolgt nun die Diskussion der strukturellen Befunde bezüglich der non-kognitiven Faktoren, sowie des SES.

5.2 Erhebung und Binnenstruktur der non-kognitiven Faktoren und des SES

Die Erfassung der non-kognitiven Faktoren bzw. des SES erfolgte mittels einer Fragebogenerhebung. Dabei ist zu beachten, dass nicht für alle 320 kognitiv getesteten Kinder auch vollständige Fragebogendaten vorliegen, da nicht alle Kinder und Eltern die Fragebögen im Anschluss an die kognitive Testung bearbeiteten (siehe Kapitel 3.3). So stellt sich die Frage, ob sich die beiden Subgruppen mit und ohne Fragebogenangaben systematisch unterscheiden, beispielsweise im Hinblick auf die kognitiven Fähigkeiten. Wäre dies der Fall, müsste der Umstand bei der Interpretation der Ergebnisse berücksichtigt werden. Der Vergleich der Mittelwerte der Summenscores für WM und Intelligenz ergibt jedoch keinen signifikanten Unterschied zwischen der Gruppe mit eingereichtem Fragebogen ($\bar{x}_{WM} = 17.69$, $\bar{x}_{CFT} = 32.90$; $N = 294$) und der Gruppe ohne Fragebogendaten ($\bar{x}_{WM} = 18.0$, $\bar{x}_{CFT} = 30.38$; $N = 26$), so dass bezüglich der kognitiven Fähigkeiten keine systematische Selektion festgestellt werden kann. Der Aspekt der Selektivität von Fragebogenerhebungen wird bei der Darstellung der Limitationen der Studie aufgegriffen werden.

Hinsichtlich der Verteilung des SES ist festzustellen, dass innerhalb der erhobenen Stichprobe keine nennenswerten Abweichungen verglichen mit einer repräsentativen Vergleichsstichprobe vorliegen (siehe Kapitel 3.3). Dies ist als positiv zu bewerten, da sich in vergleichbaren Studien, wie bspw. der Zwillingsstudie Kosmos (Spinath & Wolf, 2006), nicht selten eine Überrepräsentation höherer Schichten zeigt (Spengler, 2009), was möglicherweise mit einem gesteigerten Interesse und einer höheren Wertschätzung empirischer Forschung in bildungsnahen Familien zusammenhängt.

Zwar ergibt sich auch in der vorliegende Studie eine leichte Überrepräsentation von Personen mit Fachabitur oder Abitur (im Vergleich zur Destatis-Stichprobe), allerdings fällt diese mit einer Abweichung von weniger als 3% relativ gering aus. Außerdem ist zu beachten, dass in der Stichprobe ein deutlich größerer Anteil an Missings vorliegt, als im Vergleichsdatensatz. Somit ist unklar, ob sich die tatsächliche Verteilung im Falle vollständiger Daten als noch repräsentativer dargestellt hätte, beispielsweise weil vor allem Personen mit niedrigem Abschluss diesbezügliche Angaben verweigerten, oder ob sich eine akzentuiertere Verschiebung hin zum oberen Ende der Verteilung ergeben hätte.

Die repräsentative Zusammensetzung der untersuchten Stichprobe wirkt sich positiv auf die Generalisierbarkeit der Befunde aus. Denn eine Unterrepräsentation niedriger sozialer Schichten beeinflusst durch die resultierende Varianzeinschränkung die korrelativen Zusammenhänge zu den gleichzeitig betrachteten Prädiktoren und zum Kriterium (Spengler, 2009). Auf diese Zusammenhänge wird im Rahmen der Diskussion der Bedeutung des SES als Prädiktor schulischen Erfolgs näher eingegangen.

Neben der Verteilung des SES innerhalb der untersuchten Stichprobe, wurden die korrelativen Zusammenhänge des höchsten Bildungsabschlusses zwischen beiden Elternteilen betrachtet. Die Tatsache, dass der höchste Abschluss der Väter zu $r = .63$ mit dem höchsten Abschluss der Mütter korreliert, zeigt, dass Mütter und Väter innerhalb eines Elternpaares typischerweise ähnliche Bildungsabschlüsse aufweisen, was dem Befundmuster entspricht, dass sich Personen mit ähnlichem SES clustern (Leventhal & Brooks-Gunn, 2003).

Bezüglich der Erfassung der non-kognitiven Prädiktoren FSK und IM zeigt sich, dass alle Skalen trotz ihrer Kürze (je drei Items) Reliabilitäten im guten bis sehr guten Bereich aufweisen. Dies bedeutet, dass die Kinder die verschiedenen inhaltlichen Aspekte innerhalb eines Schulfachs ähnlich bewerten. Die Wichtigkeit der fachspezifischen Erfassung manifestiert sich im Vergleich der Reliabilitäten zwischen den Fächern, da für Mathematik ($\alpha_{FSKm} = .81$, $\alpha_{IMm} = .72$) verglichen mit Deutsch ($\alpha_{FSKd} = .65$, $\alpha_{IMd} = .65$) höhere Werte erreicht wurden. Das homogenere Antwortverhalten in Mathematik könnte auf die Tatsache zurückzuführen sein, dass die drei Items im Fach Deutsch breitere Aspekte umschreiben als die Items für Mathematik. So ist es beispielsweise denkbar, dass ein Schüler zwar gerne liest, aber weniger Interesse am Erlernen von Grammatik oder Rechtschreibregeln hat. Die Items für Mathematik beziehen sich stärker auf spezifische Rechenoperationen, für

die es wahrscheinlich weniger intraindividuell variierende Vorlieben und Fähigkeitsüberzeugungen gibt.

Die Betrachtung der Verteilungskennwerte ergibt, dass für alle Skalen relativ hohe Mittelwerte vorliegen, wodurch die für die Altersgruppe typische Zustimmungstendenz der Kinder illustriert wird. Die Mittelwerte sind jedoch mit Werten vorheriger Studien vergleichbar. Lediglich der Mittelwert von IM im Fach Mathematik ($\bar{x} = 4.04$) liegt höher als beispielsweise der in der Studie von Spinath et al. (2006) ermittelte Wert von $\bar{x} = 3.46$. Außerdem fallen die Standardabweichungen im Fach Deutsch ($SD_{FSK} = .73$, $SD_{IM} = .81$) für beide Skalen im Vergleich zu den in der Studie von Spinath et al. (2006) berichteten Werten um 1 etwas geringer aus. Doch auch bei Steinmayr und Spinath (2009) ergaben sich Standardabweichungen unterhalb des Wertes 1, so dass es insgesamt keine Hinweise auf eine auffällige Varianzeinschränkung gibt.

Die Tatsache, dass sich alle Skalen als linksschief erweisen, bedeutet, dass eine Clusterung am oberen Skalenende vorliegt. Dieses Muster spiegelt ebenfalls die Zustimmungstendenz der Kinder wider, die in Mathematik etwas ausgeprägter aufscheint als in Deutsch. Die Mehrheit der Kinder schätzt demnach die mathematischen Fähigkeiten, sowie die Freude an diesem Schulfach als hoch ein.

Auch die Betrachtung der korrelativen Zusammenhänge der Skalen liefert wichtige Erkenntnisse im Hinblick auf die Relevanz einer fachspezifischen Erfassung. Es finden sich zwischen FSK in Deutsch und Mathematik bzw. zwischen IM in beiden Fächern signifikante Korrelationen, was als Tendenz gewertet werden kann, sich in den verschiedenen Fächern als ähnlich fähig einzustufen und ähnlich viel Freude zu empfinden. Allerdings liegt die Höhe der Zusammenhänge nur in einem niedrigen bis moderaten Bereich ($r_{FSK} = .30$, $r_{IM} = .22$), so dass durchaus fachspezifische Differenzierungen stattfinden. Die Höhe dieser Korrelationen erweisen sich als äußerst konsistent zu den von Spinath et al. (2006) berichteten Werten ($r_{FSK} = .31$, $r_{IM} = .26$), während in anderen Studien keine Korrelationen auftraten (Marsh & Hau, 2004), oder gar negative Assoziationen gefunden wurden. So berichten Steinmayr und Spinath (2009) Werte von $r = -.43$ für FSK und $r = -.12$ für IM. Allerdings deutet die Tatsache, dass diese Untersuchung an Gymnasiasten durchgeführt wurde auf altersspezifische Effekte hin, was die Annahme von Shavelson, Hubner und Stanton (1976) bestätigt, dass mit zunehmenden Alter eine Ausdifferenzierung der

Bewertungen stattfindet. Hierfür spricht auch die Studie von Skaalvik und Valas (1999), welche belegt, dass die diesbezüglichen Korrelationen zwischen der dritten und achten Klasse stets negativer werden.

Dass trotz einer gewissen Einheitlichkeit im Antwortmuster fachspezifische Differenzierungen nachweisbar sind, wird außerdem durch die deutlich stärkeren Korrelationen zwischen den Skalen innerhalb einer Domäne belegt. Denn FSK und IM korrelieren zu $r = .50$ in Deutsch und zu $r = .66$ in Mathematik. Auch diese Werte entsprechen den Werten der Studie von Spinath et al. (2006) weitestgehend ($r = .56$ in Englisch, $r = .74$ in Mathematik). Dass die Korrelationen in Mathematik höher ausfallen, ist auf die höhere Reliabilität der Skalen in der mathematischen Domäne bei einer gleichzeitig leicht höheren Varianz zurückzuführen. Dies bedeutet, dass in Deutsch auf intraindividuellem Ebene weniger homogen und auf interindividueller Ebene weniger variierend geantwortet wird, was auf Skalenebene die Korrelation zwischen IM und FSK senkt.

Im Anschluss an die Betrachtung der Binnenstruktur der verschiedenen Prädiktorvariablen, erfolgt nun die Diskussion der Befunde zur Struktur des Kriteriums Schulerfolg.

5.3 Erhebung des Kriteriums Schulerfolg

Die Betrachtung der Verteilungskennwerte der Schulnoten zeigt eine gewisse Rechtsschiefe (Schiefe = .30 - .53). Somit liegt eine Clusterung am unteren Skalenende vor, was bedeutet, dass mehr Schüler bessere Leistungen erzielen, da bei Schulnoten niedrigere Werte mit besseren Leistungen korrespondieren¹².

Außerdem können die Noten in der dritten bzw. vierten Klasse als relativ stabil bezeichnet werden, was die sehr hohen Interkorrelationen belegen ($r_{\text{Deutsch}} = .85$, $r_{\text{Mathe}} = .78$). Im Fach Mathematik variieren die Noten etwas stärker als in Deutsch ($SD_{\text{Deutsch}} = .83$, $SD_{\text{Mathe}} = .92$), was bedeutet, dass in diesem Fach eher Leistungssteigerungen oder –einbußen auftreten. Über die Domänen hinweg findet

¹² Nach der Betrachtung der deskriptiven Statistiken wurde mit umkodierten Werten gearbeitet (siehe Kapitel 4.1.5).

sich zwischen den durchschnittlichen Leistungen in Deutsch und Mathematik ein Zusammenhang von $r = .73$. Dies bedeutet, dass Kinder, die gute Leistungen im Fach Deutsch zeigen, selten merklich schlechtere Leistungen in Mathematik erbringen und umgekehrt. Vor dem Hintergrund der im vorigen Abschnitt dargestellten Korrelationen von FSK in Deutsch und Mathematik, bzw. IM in beiden Fächern, bedeutet dies, dass die Kinder hinsichtlich ihrer Fähigkeiten in Deutsch und Mathematik stärker differenzieren, als es tatsächlich durch die Noten zu erwarten wäre. Dieser Befund stellt ein typisches Ergebnis der Motivationsforschung dar (Marsh, 1986) und verdeutlicht einmal mehr die Sinnhaftigkeit einer fachspezifischen Operationalisierung. Außerdem zeigt dieses Ergebnis, dass die Einschätzung des FSKs mehr ist, als eine bloße Reflexion der jeweiligen Schulnote, wie von Kritikern angeführt wird. Denn wäre dies der Fall, müsste sich der Zusammenhang zwischen FSK in Deutsch und Mathematik der Höhe der Korrelation zwischen den Noten stärker annähern. Der Zusammenhang zwischen FSK und den Schulnoten wird bei der Diskussion der relativen Bedeutung non-kognitiver Prädiktoren ausführlicher diskutiert werden.

Die bisherige Ergebnisdiskussion bezieht sich auf die vorbereitenden Analysen. Im folgenden Abschnitt werden die Befunde der zentralen Fragestellungen der vorliegenden Studie diskutiert. So erfolgt zunächst die Diskussion der Befunde bezüglich der Rolle von WM und gf als Prädiktoren schulischer Leistung (Kapitel 5.4). Im Anschluss werden die Befunde zur relativen Bedeutung kognitiver und non-kognitiver Prädiktoren diskutiert (Kapitel 5.5), bevor die Rolle des SES thematisiert wird (Kapitel 5.6).

5.4 Die Bedeutung der kognitiven Faktoren WM und gf als Prädiktoren des Schulerfolgs

Ein wesentliches Ziel dieser Arbeit besteht in der Betrachtung der relativen Bedeutung der beiden kognitiven Faktoren WM und gf hinsichtlich ihrer prädiktiven Validität für die Vorhersage des Schulerfolgs. Da sich bisherige Studien hauptsächlich auf Studenten oder ältere Stichproben beziehen (Krumm et al., 2008), oder aber auf Gruppen mit Lernschwierigkeiten (Alloway, 2009), wird eine Ergänzung der bestehenden Befunde angestrebt. Die diesbezüglichen Ergebnisse der

Kommunalitäten-, Regressions- und Strukturgleichungsanalysen werden zu diesem Zweck nun zusammenfassend dargestellt und diskutiert, wobei vorab die Diskussion der Beziehung zwischen WM und gf erfolgt.

5.4.1 Zum Zusammenhang von WM und gf

Zahlreiche Befunde dieser Studie können als Belege einer substantiellen Beziehung zwischen WM und gf gewertet werden. So findet sich auf manifester Ebene eine hoch signifikante Korrelation von $r = .45$ zwischen den Faktoren WM und gf. Auch auf latenter Ebene kann dieser Zusammenhang mit Werten zwischen $r = .69$ und $.74$ bestätigt werden (siehe Kapitel 4.2). Diese Befunde erweisen sich als konsistent zu den in Kapitel 1.2.6 vorgestellten Studien von beispielsweise Ackerman et al. (2005), Engle et al. (1999) oder Oberauer et al. (2005), welche Erwachsene untersuchten. Die vorliegende Studie bestätigt somit die Übertragbarkeit des Befundes eines starken Zusammenhangs zwischen WM und gf auf Kinder, was die Ergebnisse von Engel de Abreu et al. (2010) oder De Jong und Das-Smaal (1995) untermauert.

Trotz der stabilen Assoziation beider Faktoren sprechen die Ergebnisse der vorliegenden Studie gleichzeitig für die Betrachtung von WM und gf als separierbare Konstrukte. So fittet das einfaktorielle Messmodell, in welchem alle vier CFT-Subtests und alle drei WM-Aufgaben auf einem gemeinsamen Faktor laden, schlechter als ein Messmodell mit zwei getrennten latenten Faktoren für WM und gf (siehe 4.2.2). Dieses Befundmuster widerspricht damit der von Kyllonen und Christal (1990) postulierten Annahme, WM und gf seien isomorph. Hingegen passen die Befunde zu den Ergebnissen der Studien von Conway et al. (2003), Fry und Hale (2000) oder Alloway et al. (2004), die ebenfalls hoch assoziierte, aber dennoch distinkte Faktoren fanden.

Doch nicht nur die Beziehung der beiden kognitiven Faktoren steht im Fokus dieser Arbeit. Auch der Vergleich der prädiktiven Validität von WM und gf wurde näher beleuchtet, worauf im folgenden Abschnitt eingegangen wird.

5.4.2 Vergleich der prädiktiven Validität von WM und gf auf manifester Ebene

Wie bereits angesprochen (siehe Kapitel 1.5.1), wurde die prädiktive Validität von WM und gf für sich genommen bereits mehrfach nachgewiesen (z.B. Alloway et al., 2005; Gathercole et al., 2003, 2004; Kuncel et al., 2004). Doch die Frage nach der relativen Bedeutung beider Faktoren ist nicht hinreichend geklärt. Aufgrund der bisherigen Forschungsbefunde wurde angenommen, dass sowohl gf als auch WM in beiden Fächern signifikant zur Vorhersage beitragen können, und sowohl gemeinsame, als auch spezifische Varianzanteile aufweisen (Hypothese 1a/1b).

Mit Hilfe von Regressions- und Kommunalitätenanalysen wurde dieser Aspekt zunächst auf manifester Ebene untersucht. In diesen erweisen sich gemäß Hypothese 1b WM und gf sowohl in Deutsch als auch in Mathematik als annähernd gleich bedeutende Prädiktoren. Dies wird durch den Vergleich der Beta-Gewichte illustriert (Deutsch: $\beta_{WM} = .26$, $\beta_{gf} = .26$; Mathe: $\beta_{WM} = .31$, $\beta_{gf} = .32$). Allerdings wird im Fach Mathematik insgesamt mehr Varianz durch WM und gf aufgeklärt als in Deutsch (17% in Deutsch; 28% in Mathematik). Dies spricht für eine stärkere Assoziation der kognitiven Fähigkeiten zur mathematischen Domäne, was die Gültigkeit von Hypothese 1c belegt. Diese wird auch durch die manifesten Korrelationen zwischen WM, gf und den Schulnoten bekräftigt, welche ebenfalls in Mathematik höher ausfallen als in Deutsch. Auf den Aspekt der stärkeren Assoziation zur mathematischen Domäne wird später tiefergehend eingegangen.

Bezüglich der relativen Bedeutung beider kognitiver Prädiktoren kann aus den Ergebnissen der Regressionsanalysen geschlossen werden, dass sowohl WM als auch gf signifikant zur Vorhersage beitragen können. Dass durch die schrittweise Regression nur ein geringer zusätzlicher Beitrag über den zuerst berücksichtigten Faktor hinaus vorhergesagt werden kann (5% - 6%), deutet auf einen beachtlichen Anteil an gemeinsamer Varianz hin. Zur Differenzierung von spezifischer und gemeinsamer Varianz wurden Kommunalitätenanalysen durchgeführt. Diese bestätigen Hypothese 1a, da sich in beiden Fächern substantielle Anteile gemeinsamer Varianz manifestieren, welche mit 45% gar den jeweils stärksten Faktor darstellen. Dennoch können auch für beide Faktoren spezifische Varianzanteile nachgewiesen werden, was für die Gültigkeit von Hypothese 1b spricht. Während sich in Mathematik die spezifischen Varianzanteile von gf (28%)

und WM (27%) als sehr ausgewogen darstellen, zeigt sich hingegen in Deutsch ein etwas stärkerer Anteil an spezifischer Varianz, die auf WM zurückzuführen ist (31%), verglichen mit 24% spezifischer gf-Varianz. Möglicherweise wird dieses Muster durch die Verwendung von ausnahmslos sprachfreiem Material im CFT bedingt, während in zwei von drei WM-Aufgaben sprachliche Stimuli eingesetzt werden, bzw. Material, das verbal umkodiert werden kann. Besonders die AS-Aufgabe weist durch die Wichtigkeit von Rechtschreibprozessen augenscheinlich eine größere Nähe zum Deutschunterricht auf, als die restlichen Aufgaben und Subtests. Diese Annahme bestätigt das Muster der manifesten Korrelationen. So zeigt die AS-Aufgabe die höchste manifeste Korrelation zur Deutschnote ($r = .37$), verglichen mit den anderen Aufgaben zur Erfassung der WM-Leistung ($r = .14 - .26$) bzw. mit den CFT-Subtests ($r = .15 - .32$).

Insgesamt bekräftigen diese Befunde die bereits dargelegte starke Assoziation der beiden kognitiven Konstrukte auf manifester Ebene, sowie die Bedeutung der gemeinsamen Varianz zur Prädiktion des Schulerfolgs, was als Bestätigung der Hypothese 1a zu verstehen ist. Darüber hinaus wird die spezifische Relevanz beider Faktoren für die Vorhersage des Schulerfolgs belegt (Hypothese 1b). Denn auch wenn ein Großteil der aufgeklärten Varianz zwischen WM und gf geteilt wird, ergibt die Berücksichtigung beider Konstrukte die beste Vorhersage, da jeder Faktor auch spezifisch zur Aufklärung beitragen kann, wenngleich das Inkrement als niedrig zu bewerten ist. Auf diesen Aspekt wird im Zuge der Diskussion der strukturgleichungsanalytischen Befunde weiter eingegangen.

Da die Methode der Regressionsanalyse, wie in Kapitel 3.5.2 dargestellt, gewisse methodische Nachteile mit sich bringt, wurde die relative Bedeutung von WM und gf für die Prädiktion des Schulerfolgs auch im Rahmen latenter Modellierungen überprüft. In diesen Modellen erfolgte die Hinzunahme eines latenten SES-Faktors. Die Befunde dieser Modellierungen bezüglich der kognitiven Faktoren werden nun zusammenfassend diskutiert.

5.4.3 Vergleich der prädiktiven Validität von WM und gf auf latenter Ebene

Die Zusammenschau der Befunde der latenten Modelle 1 bis 4, mit WM, gf und SES als Prädiktoren der Noten in Deutsch und Mathematik, ergibt, dass sich WM in beiden Fächern als stärkster Prädiktor darstellt (Deutsch: $\beta_{WM} = .45$; Mathe: $\beta_{WM} = .60$). Zwar trägt auch der latente SES-Faktor signifikant zur Vorhersage bei ($\beta_{SES} = .28$), allerdings ergibt sich keine Überschneidung der Konfidenzintervalle von WM und SES, so dass der Einfluss des SES verglichen mit WM als deutlich schwächer zu bewerten ist.

Ein wichtiger Diskussionspunkt betrifft das Ergebnis, dass gf nicht signifikant zur Vorhersage beitragen kann (Modelle 1 und 2). Dieser Befund erweist sich als inkonsistent zu den Befunden der Regressions- und Kommunalitätenanalysen, welche einen gf-spezifischen Varianzanteil anzeigen. So stellt sich die Frage nach den Ursachen dieses Musters. Ein bedeutsamer Aspekt ist sicherlich das Muster der korrelativen Zusammenhänge der latenten Faktoren. Denn gf zeigt nicht nur eine deutliche Korrelation zu WM ($r_{Deutsch} = .69$; $r_{Mathe} = .74$), sondern auch zu SES ($r_{Deutsch} = .45$; $r_{Mathe} = .45$), während die Korrelation zwischen WM und SES niedriger ausfällt ($r = .20$). Die Kommunalitätenanalysen indizierten bereits, dass ein Großteil der aufgeklärten Varianz von WM und gf geteilt wird, was sich im latenten Modell in den hohen latenten Korrelationen widerspiegelt. Durch die Hinzunahme eines weiteren Prädiktors (SES), welcher stärker mit gf als mit WM assoziiert ist, verringert sich der Anteil an gf-spezifischer Varianz so weit, dass dieser nicht mehr ausreichend groß ist, um die Signifikanzgrenze zu übersteigen. Zwar erweist sich das standardisierte Beta-Gewicht von gf in Mathematik als betragsmäßig höher im Vergleich zum Beta-Gewicht des SES-Faktors ($\beta_{gf} = .23$, $\beta_{SES} = .21$). Allerdings erreicht der Wert für gf aufgrund eines höheren Schätzfehlers nicht die Signifikanzgrenze. Der höhere Schätzfehler deutet auf eine gewisse Instabilität der Schätzung hin, so dass der Einsatz einer breiteren Testbatterie bzw. eines Verfahrens mit höherer Reliabilität möglicherweise zu einem anderen Befund geführt hätte.

Der Befund, dass SES stärker mit gf als mit WM assoziiert ist, wird in Abschnitt 5.6 diskutiert.

Die Tatsache, dass ein Großteil der Varianz von WM und gf geteilt wird, dieser gemeinsame Varianzanteil in der Modellierung von Modell 1 bzw. 2 jedoch nicht in den Pfadgewichten zum Kriterium berücksichtigt wird, da diese die jeweils spezifischen Varianzanteile der einzelnen Prädiktoren abbilden, könnte zu einer Unterschätzung des Einflusses der kognitiven Fähigkeiten geführt haben. Um die Bedeutung der kognitiven Faktoren besser mit dem Einfluss des SES kontrastieren zu können, wurde folglich ein Zwei-Ebenen-Modell geschätzt, in welchem die gemeinsame Varianz von gf und WM in einem Faktor zweiter Ordnung gebündelt wurde. In Deutsch (Modell 3) verändert dies das Ergebnismuster dahingehend, dass das Pfadgewicht von SES zu Schulerfolg auf einen Wert von $\beta_{\text{SES}} = .17$ absinkt und somit der Einfluss der kognitiven Faktoren gestärkt wird. Dennoch kann der Pfad nicht ohne signifikante Verschlechterung des Modells entfernt werden. In Mathematik (Modell 4) verliert der SES jedoch stark an Einfluss ($\beta_{\text{SES}} = .12$), so dass er nicht mehr signifikant zur Vorhersage beitragen kann. Auf die fachspezifischen Effekte des SES wird in Abschnitt 5.6 näher eingegangen.

Aus den Befunden dieser Zwei-Ebenen-Modellierung ist somit abzuleiten, dass die geteilten Varianzanteile von WM und gf für die Vorhersage des Schulerfolgs einen größeren Stellenwert einnehmen, als die spezifischen. Dieser Befund passt zur von Süß (2001) formulierten Annahme, dass die Arbeitsgedächtniskapazität als „Flaschenhals der kognitiven Leistungsfähigkeit“ (S. 115) zu charakterisieren ist. Dieser Metapher folgend werden höhere kognitive Fähigkeiten durch die Arbeitsgedächtniskapazität limitiert. Die genannten Befunde bestätigen außerdem die Studie von Alloway (2009), in welcher ebenfalls gf über WM hinaus keinen inkrementellen Beitrag zur Vorhersage leisten konnte. Einschränkend ist jedoch festzustellen, dass diese Studie Kinder mit Lernproblemen untersuchte. Andere Autoren, bspw. Luo et al. (2006) oder Vock und Holling (2008) konnten hingegen einen inkrementellen Beitrag von WM über Intelligenz hinaus nachweisen. Besonders die Studie von Vock und Holling (2008) ist interessant, da in dieser ebenfalls der CFT 20-R zum Einsatz kam. Allerdings wurde WM anhand einer umfassenden Testbatterie erfasst, was eine Erklärung für die abweichenden Resultate sein könnte. Streng genommen stellen die vorliegenden Befunde jedoch keinen Widerspruch zu den Ergebnissen von Vock und Holling (2008) dar, da im Modell der kognitiven Prädiktoren gf nicht signifikant zur Vorhersage beitrug und somit beide Studien für die prädiktive Validität vom WM sprechen.

Im Hinblick auf fachspezifische Effekte ist festzustellen, dass sich diese nicht nur bezüglich des Einflusses des SES zeigen. Denn wie in Hypothese 1d spezifiziert,

manifestiert sich in allen Modellierungen der kognitiven Prädiktoren eine stärkere prädiktive Validität im Fach Mathematik. Dies drückt sich einerseits in höheren standardisierten Betagewichten aus (Deutsch: $\beta_{\text{KOG}} = .47$; Mathe: $\beta_{\text{KOG}} = .76$) und andererseits in einem insgesamt höheren Anteil an aufgeklärter Varianz. So können in Mathematik etwa 20% mehr Varianz aufgeklärt werden als in Deutsch. Der Befund einer stärkeren Beziehung der kognitiven Fähigkeiten zu Mathematik, verglichen mit der sprachlichen Domäne, erweist sich als konsistent zu vorherigen Studien (Amthauer et al., 2001; Krumm et al., 2008). Somit spielen also für das Erzielen einer guten Deutschnote andere Faktoren neben den kognitiven Fähigkeiten eine gewichtige Rolle. Die Mathematiknote wird deutlich stärker durch kognitive Fähigkeiten bedingt.

Ein Grund für die bedeutende Rolle kognitiver Fähigkeiten in Mathematik könnte die Tatsache sein, dass die Erfassung der kognitiven Fähigkeiten in dieser Studie vorrangig auf nonverbalen Materialien beruht, welche eine hohe *g*-Sättigung aufweisen. Die Anwendung breiter angelegter Instrumente, welche stärker verbale Stimuli berücksichtigen und damit sprachliche Fähigkeiten deutlicher einbeziehen, hätte die fachspezifischen Unterschiede möglicherweise reduziert. Außerdem können Mathematiknoten als objektiver im Vergleich zu Deutschnoten bezeichnet werden, da es bei der Bewertung der Richtigkeit von Mathematikaufgaben weniger Freiheitsgrade gibt, als beispielsweise bei der Bewertung eines Aufsatzes. So ist denkbar, dass der Einsatz objektiverer Erfolgskriterien, sprich Leistungstests anstelle von Schulnoten, ebenfalls die fachspezifischen Unterschiede minimiert hätte.

Die fachspezifischen Effekte werden bei der Diskussion der relativen Bedeutung kognitiver und non-kognitiver Faktoren aufgegriffen und vertieft diskutiert werden.

Zusammengefasst sprechen die Modellierungen der kognitiven Prädiktoren und des SES für die Gültigkeit der Hypothesen 1a und 1c. WM und gf stellen distinkte Faktoren dar, wobei ein Großteil der aufgeklärten Kriteriumsvarianz auf gemeinsame Varianzanteile zurückzuführen ist. Diese sind stark genug, dass bei Hinzunahme eines weiteren Faktors (z.B. SES), welcher mit den kognitiven Faktoren assoziiert ist, keine spezifische gf-Varianz mehr nachgewiesen werden kann. Aus diesem Grund kann Hypothese 1b auf latenter Ebene nicht bestätigt werden.

Allerdings gilt zu beachten, dass die Erfassung der kognitiven Fähigkeiten aufgrund gewisser Rahmenbedingungen nicht als umfassend zu bezeichnen ist. Eine breitere Messung, mit mehr als drei WM-Subtests und zusätzlichen verbalen Intelligenzskalen, hätte möglicherweise zu einer Aufwertung der spezifischen Varianzanteile der kognitiven Faktoren beigetragen und den Einfluss des SES in

Deutsch weiter verringert. Außerdem wäre es dadurch möglich, gleichzeitig spezifische und gemeinsame Varianzanteile von WM und gf als getrennte Pfade zu modellieren, wodurch sich die Gefahr der Unterschätzung der kognitiven Prädiktoren reduzieren würde. Denn die beschriebenen Modelle berücksichtigen entweder ausschließlich spezifische (Modelle 1 und 2) oder ausschließlich gemeinsame Varianzanteile (Modelle 3 und 4).

Auch die Tatsache, dass die Reliabilitäten der kognitiven Aufgaben nicht im optimalen Bereich liegen, könnte zu einer Unterschätzung der Bedeutung der kognitiven Faktoren geführt haben. Allerdings ist die in den Modellen erreichte Varianzaufklärung von 32% in Deutsch und bis zu 58% in Mathematik als ausgesprochen gut zu bezeichnen.

Neben der vergleichenden Betrachtung der kognitiven Faktoren WM und gf als Prädiktoren schulischen Erfolgs, zielt die vorliegende Studie außerdem auf die Untersuchung der relativen Bedeutung kognitiver und non-kognitiver Prädiktoren ab. Der Fokus liegt dabei auf dem Aspekt der inkrementellen Validität non-kognitiver Faktoren über die kognitiven Faktoren hinaus. Dies meint die Fähigkeit non-kognitiver Faktoren zusätzliche Varianz über die kognitiven Fähigkeiten hinaus aufzuklären. Die non-kognitiven Faktoren werden durch die etablierten motivationalen Skalen FSK und IM repräsentiert, welche die Erwartungs- und Wert-Komponente des Modells von Eccles und Wigfield (2000) widerspiegeln. Es wird angenommen, dass die non-kognitiven Faktoren in beiden Fächern einen inkrementellen Beitrag zur Varianzaufklärung leisten können, wobei die relative Bedeutung der kognitiven Faktoren die der non-kognitiven Variablen übersteigt. Die Befunde der diesbezüglichen Strukturgleichungsanalysen werden im Folgenden zusammengefasst und diskutiert. Die Diskussion des Einflusses des SES erfolgt in Kapitel 5.6.

5.5 Die relative Bedeutung kognitiver und non-kognitiver Faktoren

Bevor die Befunde der Modelle mit kognitiven und non-kognitiven Faktoren besprochen werden, wird zunächst auf das Messmodell der non-kognitiven Faktoren eingegangen.

Der Vergleich des ein- und zweifaktoriellen Messmodells bestätigt die Annahme zweier getrennter Faktoren für FSK und IM, trotz des substantiellen Zusammenhangs von $r = .57$ in Deutsch und $r = .74$ in Mathematik. Dieses Muster spiegelt die manifesten Zusammenhänge wider und erweist sich als theoriekonform (Spinath & Steinmayr, 2008). Allerdings ist zu erwähnen, dass Fehlerkorrelationen zwischen den Indikatoren für FSK und IM zugelassen werden mussten. Während das Modell ohne korrelierte Fehler aufgrund des schlechten Fits abgelehnt worden wäre, zeigt das Modell mit korrelierten Fehlertermen einen hervorragenden Fit, was als methodische Rechtfertigung dieses Vorgehens gesehen werden kann. Aufgrund der parallelen Itemformulierung der FSK- und IM-Items, welche sich auf identische Facetten des Unterrichts beziehen, wird das Modellieren korrelierter Fehlerterme außerdem inhaltlich legitimiert (Kline, 2005). Nach der Prüfung des Messmodells der motivationalen Faktoren erfolgte die Schätzung eines Prognosemodells mit kognitiven und non-kognitiven Prädiktoren. Zunächst wurden für Deutsch und Mathematik getrennte Modelle geschätzt, in denen WM, gf, FSK, IM und SES als gleichberechtigte Prädiktoren eingingen (Modelle 5 und 6). Im Anschluss erfolgte die Schätzung von Zwei-Ebenen-Modellen (Modelle 7 und 8).

Nachfolgend werden die Befunde getrennt nach Fächern diskutiert, wobei mit dem Fach Deutsch begonnen wird.

5.5.1 Die relative Bedeutung kognitiver und non-kognitiver Prädiktoren in Deutsch

Das Ergebnis der Modellierung des Prognosemodells in Deutsch weist verschiedene Besonderheiten auf. Zunächst kehrt sich das Befundmuster dahingehend um, dass nun im Gegensatz zur separaten Modellierung der kognitiven Faktoren nicht mehr gf unterhalb der Signifikanzgrenze bleibt, sondern WM. Die Ursache dieses

scheinbaren Widerspruchs ist sicherlich nur durch Beachtung der korrelativen Zusammenhänge der verschiedenen Prädiktoren zu verstehen. So korreliert in dieser Modellierung WM nicht nur mit gf ($r = .71$) und SES ($r = .22$), es liegt außerdem eine starke Verbindung zu FSK ($r = .42$) vor, dem stärksten Prädiktor des Modells. Aufgrund dieser starken Überschneidung mit allen für das Modell relevanten Variablen bleibt schließlich keine spezifische WM-Varianz in ausreichender Höhe bestehen, die für das Überschreiten der Signifikanzgrenze notwendig gewesen wäre. Dabei sollte die nicht signifikante Vorhersageleistung des WM-Faktors nicht im dem Sinne interpretiert werden, dass WM unbedeutend für die Vorhersage des Schulerfolgs ist, sondern dass die Berücksichtigung der Kommunalitäten von großer Relevanz ist. So bestätigt sich einmal mehr die starke Überschneidung von WM und gf.

Es ist anzumerken, dass im vollständigen Modell (Modell 5) ein negatives Pfadgewicht für WM auftritt ($\beta_{WM} = -.18$), was als Hinweis auf Suppressoreffekte interpretiert werden kann. In diesem Fall wird jedoch angenommen, dass es sich um ein Datenartefakt handelt, verursacht durch die hohe Korrelation zwischen WM und gf bzw. WM und FSK. Da sich das Pfadgewicht von WM außerdem als nicht signifikant erweist, wird von einer weiteren inhaltlichen Interpretation des numerischen Wertes des Pfadgewichtes abgesehen.

Weiterhin stellt sich nicht nur der latente WM-Faktor als nicht signifikanter Prädiktor dar, denn auch IM kann keine spezifische Varianz im Kriterium aufklären. Dieser Befund ist als konsistent zu vorherigen Studien zu bezeichnen, die ebenfalls eine geringere prädiktive Validität subjektiver Wertkomponenten im Vergleich zu Fähigkeitsüberzeugungen bei Grundschulern fanden (Spinath et al., 2006; Spinath et al., 2010; Steinmayr & Spinath, 2007) und bestätigt Hypothese 2c. Außerdem gehen Eccles et al. (1984) davon aus, dass sich der Einfluss von IM besonders in Entscheidungssituationen entfalten kann, wenn es beispielsweise darum geht, zwischen verschiedenen Kursen zu wählen, so dass der Mangel an spezifischer IM-Varianz sich durchaus im Einklang mit dem theoretischen Rahmenmodell befindet (siehe Kapitel 1.3.1). Weiterhin zeigt sich ein hoher Zusammenhang zwischen FSK und IM ($r = .56$), was darauf hindeutet, dass auch zwischen den non-kognitiven Faktoren ein substantieller Varianzanteil von beiden Faktoren geteilt wird, weshalb IM, ähnlich wie WM, nicht als irrelevant betrachtet werden sollte. Hierfür spricht auch die aktuelle Studie von Steinmayr und Spinath (2009), in welcher sich die in Anlehnung an Eccles operationalisierte Wert-Komponente neben FSK als

signifikanter Faktor erwies. Allerdings wurden hier keine Grundschulkinder untersucht, sondern Gymnasiasten, so dass die Bedeutung der Wert-Komponente möglicherweise in Abhängigkeit des Alters variiert und in höheren Klassenstufen an Bedeutung gewinnt. Schließlich steigt in diesen die Anzahl der Fächer an, sowie der Komplexitätsgrad des Stoffs, was besonders für Gymnasiasten zutreffend ist. Möglicherweise zeigen nun besonders die Schüler gute Leistungen, die das jeweilige Fach mögen und sich daher intensiver mit dem zu Lernenden auseinandersetzen, so dass eine tiefere Verarbeitung stattfindet (z.B. Schiefele & Schreyer, 1994). Aufgrund des steigenden Anspruchsniveaus ist es unwahrscheinlicher, dass der Stoff „nebenbei“ erlernt wird, ohne sich damit zu beschäftigen, so dass die Kinder einen Vorteil haben dürften, denen das jeweilige Fach liegt.

Korrelative Zusammenhänge zeigen sich auch zwischen den kognitiven und non-kognitiven Faktoren, was als theoriekonform zu bezeichnen ist (siehe Kapitel 1.5.2.3) und für die von Spinath et al. (2006) angenommene wechselseitige Verstärkung von Motivation und kognitiven Fähigkeiten spricht. So erscheint es schlüssig, dass ein Mehr an Fähigkeit sich positiv auf motivationale Aspekte auswirken kann und umgekehrt. Diese Mechanismen können schließlich wiederum die Leistung positiv beeinflussen.

Doch es stellt sich die Frage, weshalb WM höhere Zusammenhänge zu FSK aufweist als gf ($r_{WM} = .42$ / $r_{gf} = .20$). Möglicherweise spielt auch hier die Operationalisierung der Konstrukte eine Rolle. Denn der in der DS-Aufgabe erforderliche Umgang mit Zahlenmaterial und die in der AS-Aufgabe notwendigen Rechtschreibprozesse sind näher an die schulischen Inhalte angelehnt, als die abstrakteren Strategien, die für die Lösung des CFTs erforderlich sind, so dass sich die Einschätzung der Fähigkeit bzw. der Freude an unterrichtsbezogenen Aspekten deutlicher in den Leistungen im WM-Test widerspiegelt.

Ein wichtiges Ergebnis dieser Modellierung bezieht sich auf die relative Bedeutung kognitiver und non-kognitiver Prädiktoren. Konträr zu Hypothese 2b, die besagt, dass der Einfluss der kognitiven Faktoren den der non-kognitiven Faktoren übersteigt, stellt sich FSK wider Erwarten als stärkster Prädiktor heraus (Modell 5/7). Denn in Deutsch leistet FSK mit einem standardisierten Pfadgewicht von $\beta_{FSK} = .74$ (parsimonischstes Modell) den stärksten Beitrag zur Vorhersage. Gf fällt hingegen mit einem Wert von $\beta_{gf} = .22$ substantiell hinter die Vorhersagekraft des FSKs zurück.

Die Tatsache, dass sich die Konfidenzintervalle nicht überschneiden, verdeutlicht die Dominanz des non-kognitiven Faktors.

Es stellt sich die Frage, ob dieses unerwartete Ergebnis durch eine Unterschätzung des Einflusses der kognitiven Fähigkeiten bedingt wird, welche einen erheblichen Varianzanteil teilen und außerdem starke Korrelationen zu allen modellrelevanten Faktoren aufweisen. Um also besser zwischen kognitiven und non-kognitiven Faktoren kontrastieren zu können, wurde ein Zwei-Ebenen-Modell modelliert (Modell 7). Dieses komplexere Modell führt zwar zu einer Verschlechterung des Modellfits, allerdings liegen alle Werte mit Ausnahme des χ^2 -Tests dennoch in einem guten bis akzeptablen Bereich. Seitens des Motivationsfaktors zweiter Ordnung zeigt sich eine höhere Assoziation zum latenten FSK-Faktor als zum IM-Faktor, was die Überlegenheit der Einschätzung der eigenen Fähigkeiten gegenüber der Wertkomponente bekräftigt und damit auch die Gültigkeit von Hypothese 2c bekräftigt. Die Assoziation des Kognitionsfaktors mit WM und gf erweist sich als sehr ausgeglichen.

Bezüglich der relativen Bedeutung der verschiedenen Prädiktorgruppen bestätigt sich das Muster, dass der Motivationsfaktor mehr prädiktive Validität besitzt als der Kognitionsfaktor ($\beta_{\text{MOT}} = .72$, $\beta_{\text{KOG}} = .24$). Die Dominanz des Motivationsfaktors zeigt sich darin, dass sich die Konfidenzintervalle des Kognitions- und Motivationsfaktors nicht überschneiden.

Die Tatsache, dass motivationale Faktoren inkrementell zur Vorhersage des Schulerfolgs beitragen können, wurde bereits in anderen Studien aufgezeigt (beispielsweise Chamorro-Premuzic et al., 2010; Greven et al., 2009; Spinath & Spinath, 2005; Spinath et al., 2006) und kann als Bestätigung der Hypothese 2a betrachtet werden. Dass jedoch das Inkrement die prädiktive Validität der kognitiven Fähigkeiten übersteigt, erweist sich als unerwartetes Ergebnis (siehe Hypothese 2b), da in bisherigen Studien die kognitiven Faktoren überwiegend die validesten Prädiktoren darstellten (z.B. Spinath et al., 2006). Die Stabilität des Befundes auch im Zwei-Ebenen-Modell spricht dafür, dass die Unterschätzung der kognitiven Fähigkeiten nur bedingt als Erklärung der Stärke des FSKs herangezogen werden kann. Als weiteren Grund wäre auch hier die geringere Reliabilität der kognitiven Skalen zu nennen. Andererseits konnten auch Steinmayr und Spinath (2009) FSK als stärksten Prädiktor identifizieren. Wie bereits erwähnt, wurden in dieser Studie zwar keine Grundschüler untersucht und es wurde keine latente Modellierung durchgeführt, allerdings wurden ebenfalls Noten als Kriterium eingesetzt, was die

Relevanz der Art der Operationalisierung des Erfolgskriteriums verdeutlicht. So konnte gezeigt werden, dass die Beziehung zwischen Noten und FSK und IM stärker ist, als zwischen Leistungstestergebnissen und den motivationalen Faktoren (Gottfried, 1990; Schicke und Fagan, 1994). Dies erklärt sich durch die Tatsache, dass Noten nie reine Leistungsindikatoren darstellen, sondern stets auch Verhaltenseigenschaften des Schülers und andere Merkmale in die Bewertung einfließen (siehe Kapitel 1.5). Somit kann angenommen werden, dass auch in der vorliegenden Studie der Einsatz von Schulnoten zur hohen prädiktiven Validität des FSKs beigetragen hat. Dies könnte besonders in Deutsch eine Rolle spielen, da wie bereits angesprochen, dem Lehrer in diesem Fach eher die Möglichkeit gegeben ist, verhaltensbezogene Aspekte mit in die Bewertung einfließen zu lassen. Es wäre beispielsweise denkbar, dass ein Kind, das sich selbst als fähiger einschätzt, aktiver am Unterricht teilnimmt und damit die Einstellung des Lehrers positiv beeinflusst.

Als Argument gegen die prädiktive Validität des Konstrukts FSK, wird üblicherweise angeführt, dass die enge Verknüpfung von FSK und Note lediglich darauf zurückzuführen ist, dass Kinder über internalisierte Repräsentationen ihrer Schulnoten verfügen, welche sie bei der Einschätzung ihrer Fähigkeiten abrufen. Damit wäre die Beantwortung der FSK-Items lediglich ein Maß der Genauigkeit des Wissens über eigene Leistungen. Diese Sichtweise entspricht dem skill-development-Ansatz (Marsh, Byrne & Yeung, 1999), welcher besagt, dass das aktuell wirksame FSK durch die vorherigen Leistungen bedingt wird und nicht umgekehrt. Selbstverständlich ist es zutreffend, dass die Vorleistungen das aktuelle FSK beeinflussen, doch es gilt als empirisch abgesichert, dass auch umgekehrte Zusammenhänge bestehen, so dass insgesamt von reziproken Effekten auszugehen ist. Diese konnte mehrfach in längsschnittlichen Studien nachgewiesen werden (z.B. Guay et al., 2003; Chamorro-Premuzic et al., 2010). Außerdem konnten Steinmayr und Spinath (2009) selbst bei Kontrolle der Vorleistung einen inkrementellen Beitrag motivationaler Komponenten über Intelligenz hinaus nachweisen. Schließlich ist dieser Kritik der bereits erwähnte Sachverhalt entgegenzusetzen, dass die Noten in Deutsch und Mathematik höher korrelieren, als das FSK, was der Idee widerspricht, dass die Kinder sich bei ihren Einschätzungen lediglich ihre aktuellen Noten ins Gedächtnis rufen.

Die Tatsache, dass die Kinder zwischen ihren Fähigkeiten in Deutsch und Mathematik stärker differenzieren als die Noten indizieren, spricht außerdem für die Gültigkeit des I/E Frame of reference model (Marsh, 1986), wonach nicht nur externale Vergleichsmaßstäbe die Ausprägung des FSKs bedingen, sondern auch

internale Prozesse. Dass ein Schüler in beiden Fächern etwa gleich gute Leistungen erbringt, heißt nicht, dass er sich nicht in einem Fach stärker anstrengen muss, um gleich gute Zensuren zu erzielen. Diese subjektiv wahrgenommenen Unterschiede können zu einer differenzierten Bewertung der eigenen Fähigkeiten in beiden Fächern beitragen.

Bevor die relative Bedeutung kognitiver und non-kognitive Faktoren weiter diskutiert wird, erfolgt zunächst die Diskussion der Befunde im Fach Mathematik.

5.5.2 Die relative Bedeutung kognitiver und non-kognitiver Prädiktoren in Mathematik

Das parsimonischste Prognosemodell für Mathematik mit den gleichberechtigten Prädiktoren WM, gf, FSK, IM und SES zeigt ebenfalls wie in Deutsch einen hervorragenden Gesamtfit (Modell 6f). Analog zu Deutsch bleiben IM und WM unterhalb der Signifikanzgrenze. Mögliche Gründe hierfür wurden bereits angesprochen.

Darüber hinaus zeigen gf und FSK nahezu identische Beta-Gewichte und Konfidenzintervalle ($\beta_{FSK} = .58$, $\beta_{gf} = .57$), so dass sich die Dominanz des FSK-Faktors in diesem Fach nicht bestätigt, wenngleich der relative Einfluss von gf dem des FSKs entspricht.

Im Zwei-Ebenen-Modell (Modell 8a), in welchem die gemeinsame Varianz von WM und gf gebündelt wird, ebenso wie die gemeinsame Varianz von FSK und IM, reduziert sich der Einfluss des FSKs schließlich weiter, so dass sich der Kognitionsfaktor als stärkster Prädiktor durchsetzen kann ($\beta_{KOG} = .62$, $\beta_{MOT} = .43$). Zwar kann der Motivationsfaktor nach wie vor inkrementell zur Vorhersage beitragen, doch die Berücksichtigung des gemeinsamen Varianzanteils von WM und gf bei der Vorhersage der Mathematiknote führt zu dem Befund, dass die relative Bedeutung der motivationalen Faktoren hinter die der kognitiven Faktoren zurückfällt. Die Überlappung der Konfidenzintervalle zeigt jedoch, dass der Einfluss des Motivationsfaktors auch hier nicht zu vernachlässigen ist. Dennoch erweisen sich in Mathematik die kognitiven Faktoren insgesamt erwartungskonform als mächtigste Einflussgrößen (Hypothese 2b).

Damit bestätigt sich erneut die Wichtigkeit der Bewertung der relativen Bedeutung der Prädiktorgruppen auf fachspezifischer Ebene (siehe Hypothese 2d), da sich die kognitiven Faktoren als besonders prädiktiv bezüglich der Mathematiknote erweisen. Die relative Bedeutung entspricht hier der Bedeutung der motivationalen Faktoren. In Deutsch konnte jedoch ein stärkerer Einfluss der motivationalen Faktoren gegenüber den kognitiven Faktoren nachgewiesen werden. Auch in den Studien von Eccles (1984) und Spinath et al. (2006) konnte eine stärkere Assoziation motivationaler Variablen zur sprachlichen Domäne festgestellt werden, während sich bei Steinmayr und Spinath (2009) ein umgekehrtes Ergebnismuster zeigte. Das abweichende Muster der letztgenannten Studie könnte möglicherweise auf eine Varianzeinschränkung in den kognitiven Maßen zurückzuführen sein, da ausschließlich Gymnasiasten untersucht wurden.

Allgemein sind die Gründe für diese fachspezifischen Effekte sicherlich vielschichtig. Ein bereits diskutierter Aspekt ist die unterschiedliche Objektivität der Noten in Deutsch und Mathematik, wodurch in Deutsch Verhaltensaspekte wie aktive Teilnahme am Unterricht, Generieren von eigenen Ideen, Kreativität, Betragen, Fleiß oder Aufmerksamkeit leichter belohnt werden können bzw. eher in die Notengebung einfließen können. Selbstverständlich werden in Deutsch auch Facetten wie Korrektheit von Grammatik und Rechtschreibung bewertet, welche ebenso eindeutig zu klassifizieren sind, wie die Ergebnisse von Rechenaufgaben. Dennoch ergeben sich beispielsweise beim Schreiben von Aufsätzen vermehrt Möglichkeiten, kreative Ideen oder die Anstrengungsbereitschaft des Kindes zu berücksichtigen, ein Umstand, der in Mathematik so ausgeschlossen ist. Damit wäre es in Deutsch leichter, mangelnde kognitive Fähigkeit bis zu einem gewissen Grade durch Motivation zu kompensieren. Dass daraus nicht zu schließen ist, dass die Deutschnoten unreliable und fähigkeitsferne Bewertungen darstellen, zeigt die deutliche Korrelation zwischen den Deutsch- und Mathematiknoten ($r = .74$). Denn dies bedeutet, dass es zwar Notenunterschiede gibt, in der Tendenz ein guter Deutschschüler aber auch gute Leistungen in Mathematik erbringt und umgekehrt.

Neben der Bewertungssituation in Deutsch und Mathematik, hat vermutlich auch die Formulierung der Items für FSK und IM zu den fachspezifischen Unterschieden beigetragen. So reflektieren die drei Items für Deutsch sehr präzise die wichtigsten Unterrichtsinhalte Rechtschreibung, Lesen und Schreiben von Aufsätzen. Zwar beziehen sich auch die drei mathematischen Aspekte Multiplizieren/Dividieren, Kopfrechnen und Textaufgaben auf wichtige Fähigkeiten, die im Mathematikunterricht

erforderlich sind. Möglicherweise sind diese drei Facetten jedoch weniger direkte Indikatoren der Mathematiknote, weshalb der Zusammenhang geringer ausfällt.

Wie bereits in Kapitel 1.5.1 erwähnt, ist anzunehmen, dass auch die Wahl der kognitiven Tests zu einer Verstärkung der fachspezifischen Effekte beigetragen hat. So war zu erwarten, dass die Lösung des CFTs, welcher eine hohe *g*-Sättigung aufweist, besonders deutlich mit den mathematischen Leistungen assoziiert ist, da schließlich keinerlei verbales Material zum Einsatz kam. Der Einsatz eines Intelligenztests mit zusätzlichen verbalen Skalen hätte möglicherweise die Bedeutung der Intelligenz in Deutsch gesteigert. Zwar wurden im Arbeitsgedächtnistest auch verbale Materialien eingesetzt, doch auch hier hätte eine breitere Messung möglicherweise zu einer Steigerung des Einflusses in Deutsch geführt. Ein Indiz hierfür ist die Tatsache, dass die AS-Task unter allen kognitiven Subtests die höchste Korrelation zur Deutschnote aufweist.

Außerdem ist zu bemerken, dass für die kognitiven Tests höhere Reliabilitäten erstrebenswert wären. Dass das Befundmuster jedoch nicht vollständig auf die Reliabilitätsunterschiede der kognitiven und non-kognitiven Skalen zurückzuführen ist, zeigt die Tatsache, dass die kognitiven Fähigkeiten in Mathematik mehr Einfluss als die motivationalen Faktoren ausüben (bzw. gleich viel), obwohl die Reliabilität der Motivationsskalen in Mathematik höher ausfällt als in Deutsch.

Die fachspezifischen Effekte stellen außerdem ein weiteres Argument gegen die bereits thematisierte Ansicht dar, Fähigkeitsüberzeugungen seien lediglich eine Wiedergabe des internalisierten Feedbacks bezüglich vorheriger Leistungen. Schließlich ist nicht ersichtlich, weshalb Schüler ihre Schulnoten in Deutsch besser internalisieren sollten als in Mathematik.

Zuletzt sei noch einmal auf die Rolle der Operationalisierung des Kriteriums verwiesen. Das Einsetzen objektiverer Schulleistungstests hätte möglicherweise sowohl den Einfluss der motivationalen Faktoren insgesamt geschwächt, als auch die fachspezifischen Effekte reduziert. Dennoch scheint die Wahl von Schulnoten als Erfolgskriterium in mehrfacher Hinsicht gerechtfertigt. Zum einen sind Noten zwar allgemein höher mit motivationalen Maßen assoziiert (Gottfried, 1990), aber dennoch stellen sie ein wichtiges „real life“-Kriterium dar, da sie im Gegensatz zu Leistungstests den Schulalltag der Kinder deutlich prägen (Steinmayr & Spinath, 2009). Außerdem scheinen sie als Schlüssel für Karrieremöglichkeiten zu fungieren

und eignen sich damit sehr gut als Indikatoren für akademischen und beruflichen Erfolg (Roth & Clark, 1998), weshalb sie das wichtigere Kriterium in Bezug auf schulische Erfolge darstellen. Zum anderen korrelieren Noten hoch mit Leistungstestergebnissen (Hoge & Coladarci, 1989), was zeigt, dass die beiden Formen der Kriteriumsoperationalisierung nicht völlig Verschiedenes erfassen. Zuletzt ist die besonders für das deutsche Schulsystem charakteristische Relevanz der Noten zu verweisen. Schließlich basiert die Entscheidung, welche Schule nach der Grundschule besucht werden soll, besonders stark auf den Zensuren der vierten Klasse. Allerdings könnte sich die Bedeutung der Noten der vierten Klasse durch die Einführung der Gemeinschaftsschule abschwächen, da durch die Zusammenlegung von Real- und Hauptschule weniger stark selektiert wird, bzw. weniger früh Gruppen mit unterschiedlichem Leistungsniveau separiert werden.

Neben der Betrachtung der relativen Bedeutung kognitiver und non-kognitiver Prädiktoren für die Vorhersage der Deutsch- und Mathematiknoten, zielte die vorliegende Studie außerdem darauf ab, den Einfluss des SES zu beleuchten. Diesbezügliche Befunde werden im folgenden Abschnitt thematisiert.

5.6 Zur relativen Bedeutung des SES als Prädiktor des Schulerfolgs

In allen Strukturgleichungsanalysen wurde ein latenter SES-Faktor modelliert, der den höchsten Bildungsabschluss der Eltern repräsentierte. Es wurde betrachtet, ob der elterliche Bildungshintergrund, wie oft diskutiert, über die kognitiven Fähigkeiten hinaus signifikant zur Vorhersage beitragen kann. Außerdem wurde der Zusammenhang zwischen den kognitiven Fähigkeiten und SES untersucht, da angenommen wurde, dass es zwar signifikante Korrelationen zwischen SES und sowohl WM als auch gf gibt. Allerdings wurde eine stärkere Beziehung zu gf als zu WM erwartet.

Dieser in Hypothese 3a spezifizierte Sachverhalt zeigt sich in allen Modellen. Die Tatsache, dass SES stärker mit gf als mit WM assoziiert ist, zeichnet sich bereits bei der Betrachtung der manifesten Korrelationen ab ($r_{WM} = .15$; $r_{gf} = .26$). Dieser Befund bestätigt das in Kapitel 1.4 thematisierte Verständnis von WM als wenig durch

Vorerfahrung oder sozioökonomische Umweltbedingungen beeinflussbares Konstrukt, welches das Lernpotenzial eines Kindes relativ unverfälscht indizieren kann (Alloway, 2009, Alloway et al., 2004). Dass dennoch signifikante Korrelationen nachzuweisen sind, ist vermutlich darauf zurückzuführen, dass zwei der drei eingesetzten WM-Aufgaben die Anwendung gelernter Kulturtechniken erfordern (Schreiben von Zahlen und Wörtern), was in vielen computerbasierten WM-Aufgaben nicht der Fall ist.

Weiterhin zeigt die moderate Korrelation zwischen gf und SES, dass sich die Ergebnisse im CFT nicht als gänzlich unabhängig von der sozialen Herkunft bzw. vom Bildungshintergrund erweisen, auch wenn der CFT mit dem Ziel einer möglichst geringen Statusabhängigkeit konzipiert wurde. Die Höhe der Korrelation zwischen gf und SES liegt jedoch etwas unterhalb des in der Literatur für den Zusammenhang zwischen Intelligenz und SES berichteten durchschnittlichen Wertes von $r = .30 - .40$ (Brody, 1992), was dennoch die Annahme, dass der CFT sich als weniger statusabhängig erweist als andere Verfahren, bekräftigt.

Allgemein gilt hier zu berücksichtigen, dass die Korrelation zwischen gf und SES teilweise auf genetische Effekte zurückzuführen ist, da sich die elterliche Intelligenz, welche sich im erzielten Bildungsabschluss niederschlägt, an die Kinder weiter vererbt (Bouchard & McGue, 2003), so dass es schlüssig ist, dass ein Zusammenhang zwischen der Leistung der Kinder in kognitiven Tests und dem höchsten Bildungsabschluss der Eltern besteht (siehe Kapitel 1.4).

Die Unterschiede der Zusammenhänge zu WM und gf können im Rahmen dieser Studie nicht auf spezifische Prozesse zurückgeführt werden, da die Erfassung der Konstrukte zu unspezifisch und zu wenig umfassend ist. Auch die Operationalisierung des latenten SES-Faktors anhand des höchsten Bildungsabschlusses der Eltern ist als knapp zu bezeichnen. Denn dieser stellt neben Einkommen und Beruf lediglich einen der drei wichtigsten Indikatoren des familiären SES dar (Ensminger & Fothergill, 2003; siehe Kapitel 1.4). Die Aufteilung in fünf Kategorien ist desweiteren als zu grob zu klassifizieren, um differenzierte Unterschiede der sozialen Schichtzugehörigkeit anzuzeigen. Bei der Interpretation des Einflusses des SES-Faktors sollte daher stets berücksichtigt werden, dass eine breitere Messung unter Verwendung weiterer Indikatoren möglicherweise andere Ergebnisse geliefert hätte. Die Erfassung des Einkommens wäre in diesem Zusammenhang vorteilhaft, da es sich unkompliziert klassifizieren und hierarchisch ordnen lässt. Allerdings besteht die Gefahr, dass die Frage nach dem Einkommen als zu persönlich empfunden wird, woraus ein hoher Datenverlust resultieren kann.

Hierdurch wird wiederum die Anwendbarkeit der Variable zur Datenauswertung eingeschränkt (Spengler, 2009).

Hinsichtlich der prädiktiven Validität des SES-Faktors ergeben sich fachspezifische Effekte. Während der höchste Bildungsabschluss der Eltern in Deutsch stets signifikant zur Vorhersage beitragen kann, erweist sich der Einfluss in Mathematik als vernachlässigbar bzw. als nicht signifikant. Somit kann Hypothese 3b nicht bzw. nur teilweise bestätigt werden. Die standardisierten Beta-Gewichte variieren in Deutsch zwischen $\beta_{SES} = .17$ und $\beta_{SES} = .28$, so dass der Einfluss des SES verglichen mit FSK als gering zu bewerten ist. Allerdings überschneidet sich das Konfidenzintervall des SES erheblich mit dem Intervall des Kognitionsfaktors im Zwei-Ebenen-Modell, so dass die prädiktive Validität der beiden Faktoren in Deutsch vergleichbar ist.

In Mathematik trägt der SES lediglich dann signifikant zur Vorhersage bei, wenn WM und gf als konkurrierende Prädiktoren modelliert werden, ohne Hinzunahme der motivationalen Faktoren. Auch im Zwei-Ebenen-Modell der kognitiven Faktoren kann der Pfad von SES zu Schulerfolg auf null fixiert werden. Die Ergebnisse in Mathematik bestätigen somit die von Colom und Flores-Mendoza (2007) vertretene Auffassung, dass der Einfluss des SES auf Schulleistung zu vernachlässigen ist, während die Befunde für Deutsch Studien stützen, die für die inkrementelle Validität des SES sprechen (Schick et al., 2006).

Dass gerade in Deutschland familiäre Merkmale wie SES und Disparitäten der Bildungsbeteiligung eng verwoben sind, wurde bereits im Rahmen der PISA-Untersuchung (OECD, 2001) und anderen großangelegten Studien aufgezeigt (Baumert et al., 2003; Müller-Benedict, 2007). Doch dies erklärt nicht, weshalb die prädiktive Validität in der vorliegenden Studie lediglich im Fach Deutsch aufscheint.

Auch der fachspezifische Einfluss des SES wird vermutlich durch die bereits thematisierte geringere Objektivität der Deutschnoten begünstigt, welche bei der Bewertung Raum für nicht-fähigkeitsbezogene Aspekte lassen, so dass sich die Frage stellt, ob objektivere Bewertungsformen im Fach Deutsch anzustreben wären, wenn es darum geht Bildungsdisparitäten zu verringern. Allerdings könnte auch der Umstand eine Rolle spielen, dass Kinder in bildungsferneren Familien weniger Kontakt zu Büchern und anderen Lernmedien haben. Es ist anzunehmen, dass sich Eltern in bildungsferneren Familien seltener mit Büchern oder anderen Medien bildungsfördernd beschäftigen, so dass sie weniger als Vorbilder dienen können. Dadurch könnte sich insgesamt eine mangelnde Erfahrung im Umgang mit Texten

und Ähnlichem einstellen, was vor allem im Grundschulalter negative Konsequenzen für die Entwicklung sprachlicher Fähigkeiten und damit für die Leistung in Deutsch haben könnte (Jeynes, 2002b; Petrill et al., 2004). So fanden auch Oswald und Krappmann (2004) Belege für die Bedeutung der Anzahl der Bücher in einem Haushalt als Indikator des Anregungsgehalts der familiären Umgebung im Hinblick auf die Vorhersagbarkeit von Schulnoten. Selbstverständlich ist eine Beziehung zwischen der elterlichen Intelligenz und dem Anregungsgehalt der familiären Umwelt zu vermuten, so dass auch hier eine genetische Komponente der Beziehung zu den Schulnoten zugrunde liegt. Die Tatsache, dass jedoch fachspezifische Effekte des SES vorliegen, spricht insgesamt dafür, dass der Einfluss des SES nicht vollständig durch die genetische Basis, die sich Intelligenz und SES teilen, erklärt werden kann. So sind die Zusammenhänge aufgrund der Kovariation bzw. Interaktion von Genetik und Umwelt nie ausschließlich erbpsychologisch oder ausschließlich umweltbedingt zu interpretieren.

Möglicherweise wirkt sich in Deutsch die Einstellung des Lehrers gegenüber dem Schüler und das Wissen um den sozialen Hintergrund des Schülers unbewusst auf das Leistungsurteil derart aus, dass Schüler aus Familien mit höherem SES tendenziell besser bewertet werden als Schüler aus Familien mit niedrigerem SES. Der objektivere Bewertungsprozess in Mathematik hingegen könnte einem derartigen Bewertungsbias entgegenwirken. Dieser Aspekt passt zu den bereits in der internationalen IGLU-Studie 2006 (Internationale Grundschul-Lese-Untersuchung, 2006) aufgedeckten Befunden, dass Kinder aus bildungsfernen Familien deutlich höhere Testleistungen als Akademikerkinder erbringen müssen, um von ihren Lehrern eine Empfehlung für das Gymnasium zu erhalten. So war bei gleicher kognitiver Fähigkeit und Leseleistung, die Chance einer Gymnasialempfehlung bei Kindern von Facharbeitern oder Angestellten mehr als zweieinhalb mal geringer als bei Kindern aus Familien mit Eltern, die der oberen Dienstklasse angehörten (vgl. Bos et al., 2008).

Es wäre außerdem denkbar, dass die Einstellung des Lehrers im Sinne einer sich selbst erfüllenden Prophezeiung wirkt, in dem der Schüler durch das Verhalten des Lehrers spürt, wenn er für weniger fähig eingeschätzt wird, was sich schließlich in der eigenen Einschätzung und den Leistungen widerspiegeln könnte (Jussim, Robustelli & Cain, 2009).

Es stellt sich die Frage, ob sich die Einführung der Gemeinschaftsschule dahingehend positiv auswirken kann, dass durch das längere gemeinsame Lernen einer Stereotypisierung oder Stigmatisierung, beispielsweise im Sinne eines negativ besetzten Labels wie *Hauptschüler*, entgegengewirkt wird. Schließlich gilt es als ein pädagogisches Ziel dieser Schulform, „das gegenseitige Verstehen, die Bereitschaft zu sozialem Handeln und die Fähigkeit zur Zusammenarbeit von Menschen unterschiedlicher außerschulischer Lernvoraussetzungen und sozialer Herkunft“ (Saarländisches Ministerium für Bildung, 2011, S.1) zu fördern.

Im Anschluss erfolgt nun eine kurze Zusammenfassung der wichtigsten Kernaspekte der vorliegenden Arbeit.

5.7 Fazit

Die vorliegende Untersuchung verfolgte dreierlei Ziele. Einerseits wurden die kognitiven Faktoren WM und gf hinsichtlich ihrer prädiktiven Validität in den Fächern Deutsch und Mathematik verglichen. Darüber hinaus wurde geprüft, inwiefern non-kognitive Faktoren (FSK, IM) inkrementell zur Vorhersage beitragen können. Schließlich wurde die Bedeutung des höchsten Bildungsabschlusses der Eltern als Indikator des SES im Hinblick auf die Schulerfolgsprognose betrachtet.

Fasst man nun die wichtigsten Befunde zusammen, zeigt sich, dass sowohl kognitive als auch non-kognitive Faktoren zur Vorhersage des Schulerfolgs in beiden Fächern beitragen. Die relative Bedeutung variiert je nach Schulfach. Während in Deutsch die prädiktive Validität der non-kognitiven Faktoren die Validität der kognitiven Faktoren übersteigt, gleicht sich in Mathematik die Vorhersagekraft beider Prädiktorgruppen an. Fokussiert man die gemeinsame Varianz von WM und gf, kann dieser Varianzanteil in Mathematik schließlich stärker zur Vorhersage beitragen als die non-kognitiven Faktoren.

Somit zeigt sich, dass besonders der geteilte Varianzanteil der kognitiven Faktoren prädiktive Validität besitzt, während die spezifischen Anteile eine untergeordnete Rolle spielen. Allerdings bleibt offen, ob dies der Tatsache geschuldet ist, dass nicht gleichzeitig gemeinsame und spezifische Varianzanteile modelliert werden konnten. Aufgrund der Befunde der Kommunalitätenanalysen ist jedoch in Frage zu stellen, ob sich das Befundmuster bei einer derartigen Analyse anders dargestellt hätte. Allerdings könnte die wenig breit gefasste Konstruktoperationalisierung zu einer Unterschätzung der spezifischen Varianzanteile geführt haben.

Die Frage, ob nun WM oder gf den valideren Prädiktor der Schulleistung darstellt, kann aufgrund des inkonsistenten Befundmusters nicht abschließend beantwortet werden. Grundsätzlich sind beide Faktoren zur Prädiktion der Schulleistung geeignet. Ein Vorteil der Messung des WMs ist jedoch, dass spezifische Lernschwierigkeiten und –defizite aufgedeckt werden können, da auf Prozessebene verschiedene Komponenten differenzierbar sind, die mit spezifischen Lernproblemen in Verbindung stehen (Carretti et al., 2009). Da diese Defizite nicht zwangsläufig durch mangelnde Intelligenz verursacht werden, ist der Einsatz der Arbeitsgedächtnismessung als diagnostisches Instrument von äußerster praktischer Relevanz (Carretti et al., 2009).

Vorteilhaft ist außerdem die geringere Korrelation zu Indikatoren des SES (z.B. Campbell et al., 1997; Engel et al., 2008).

Der Nachweis der prädiktiven Validität motivationaler Faktoren widerspricht den Befunden von Gagné und St Père (2001), die den Einfluss non-kognitiver Faktoren anzweifeln und bestätigt im Gegenzug beispielsweise die Studien von Spinath et al. (2006), Steinmayr und Spinath (2009) oder Schicke und Fagan (1990). Die herausragende prädiktive Validität des FSKs in Deutsch entspricht jedoch nicht den Erwartungen. Dennoch zeigen sich auch in anderen Studien vergleichbare Befunde (Steinmayr & Spinath, 2009). Die Tatsache, dass IM nicht signifikant zur Vorhersage beitragen kann, ist als erwartungskonform zu bezeichnen, wobei offen bleibt, ob es sich hierbei um ein altersspezifisches Muster handelt, das möglicherweise mit steigender Klassenstufe Änderung erfährt (Steinmayr & Spinath, 2009).

Der Einfluss des SES erweist sich in der vorliegenden Studie als fachspezifisch. In Deutsch kann zwar ein spezifischer Varianzanteil durch SES aufgeklärt werden, der im Zwei-Ebenen-Modell dem Einfluss des Kognitionsfaktors entspricht. Dieser erweist sich jedoch als deutlich geringer, verglichen mit dem Einfluss des FSKs. In Mathematik verliert der SES schließlich vollständig an Vorhersagekraft. Die Gründe des fachspezifischen Ergebnisses gehen teilweise auf die Operationalisierung des Kriteriums zurück.

Im Anschluss erfolgt die kritische Betrachtung der Grenzen dieser Studie, sowie ein Ausblick auf weitere Forschungsfragen.

5.8 Einschränkungen und Grenzen der Studie

Bezüglich der Kritikpunkte dieser Studie ist zunächst anzumerken, dass es sich um eine querschnittliche Untersuchung handelt. Dies bedeutet, dass die Datenerhebung eine Momentaufnahme darstellte, die streng genommen kausale Interpretationen der gefundenen Zusammenhänge im Sinne des Ursache-Wirkungs-Prinzips, erschwert. Eine längsschnittliche Untersuchung hätte die Frage, inwiefern vorherige Fähigkeitsüberzeugungen nachfolgende Noten besser als kognitive Fähigkeiten vorhersagen können, eher ermöglicht. Außerdem hätten die jeweiligen Vorleistungen kontrolliert werden können. So ist die gleichzeitige Erhebung von Prädiktor und

Kriterium nur im weitesten Sinne als Prädiktion zu bezeichnen, besonders vor dem Hintergrund, dass auch Noten der dritten Klassenstufe einbezogen werden, was bedeutet, dass auch Kriteriumsfacetten einfließen, die bereits vor Erhebung der Prädiktoren vorlagen. Die beachtliche Stabilität der Noten (Grube & Hasselhorn, 2006; Tent, 2006) rechtfertigt jedoch dieses Vorgehen und lässt es als unwahrscheinlich erscheinen, dass durch diese Operationalisierung eine substantielle Ergebnisverzerrung verursacht wurde. Ein weiterer Grund für die querschnittliche Erhebung der Konstrukte ist die Tatsache, dass die Untersuchung einer so jungen Schülerstichprobe an viel Einsatz und Hilfsbereitschaft der jeweiligen Lehrkräfte und Eltern gebunden ist, so dass darauf zu achten war, dass die Erhebung mit möglichst geringem Aufwand für Schule und Eltern ablaufen konnte. Eine längsschnittliche Erfassung der Noten hätte diesen Aufwand deutlich erhöht und zusätzlich die Gefahr gesteigert, dass nicht alle Noten ein halbes Jahr nach der eigentlichen Datenerhebung weitergegeben worden wären, wodurch sich die Stichprobe womöglich stark verringert hätte. Die querschnittliche Erfassung wurde daher in Folge der Kosten-Nutzen-Abwägung präferiert.

Die Wahl der Schulnoten als Kriterium des Schulerfolgs wurde bereits mehrfach angesprochen. Als Kritik am Einsatz von Schulnoten als Leistungskriterium wird häufig angeführt, dass keine verbindlichen Richtlinien existieren, nach denen Noten vergeben werden. So orientieren sich Lehrer häufig am klassenspezifischen Bezugsrahmen, was die Vergleichbarkeit von Noten erschwert (Tent, 1998). Die Tatsache, dass hohe Zusammenhänge zwischen Noten und Leistungstestergebnissen nachgewiesen werden konnten (Hoge & Coladarci, 1989) bescheinigt den Lehrern im Allgemeinen jedoch gute diagnostische Fähigkeiten. Auch die in der vorliegenden Studie berichteten Korrelationen der verschiedenen Prädiktoren zu den Noten, bzw. der Noten untereinander, erweisen sich durchweg als theoriekonform, so dass der Einsatz der Noten als schulrelevantes Kriterium gerechtfertigt ist.

Ein weiterer Diskussionspunkt, der die Generalisierbarkeit der Befunde betrifft, bezieht sich auf das relativ junge Alter der Stichprobe. Wie in Kapitel 1.3.2 dargestellt, bewerten Grundschüler ihre Fähigkeiten und Interessen noch undifferenzierter und positiver als dies ältere Schüler tun. Damit sind die Ergebnisse dieser Studie nicht ohne weiteres auf ältere Schüler übertragbar. Da der Institution Grundschule jedoch durch die Vermittlung wichtiger Basiskompetenzen, die für die weitere schulische Laufbahn entscheidend sind, eine enorme Verantwortung

zukommt, ist es von großer Relevanz zu untersuchen, welche Faktoren zu guten Leistungen in der Grundschulzeit beitragen. Nur so ist die sinnvolle Entwicklung von Interventions- und Fördermöglichkeiten bzw. das Erkennen und Ausschöpfen des jeweiligen Lern- und Leistungspotenzials möglich.

Auch die Generalisierbarkeit der Befunde zur Erfassung der kognitiven Fähigkeiten ist zu diskutieren, da die kognitive Testung an gewisse Vorgaben gebunden war, die die Messung beeinflusst haben könnten. So erfolgte die Erfassung der Arbeitsgedächtniskapazität nicht durch ein standardisiertes Verfahren. Vielmehr wurden gängige Aufgaben auf die Testsituation im Schulsetting angepasst, was zu einer Einschränkung der Validität geführt haben könnte. Zwar wurde durch die Durchführung von Pilotstudien die Anwendbarkeit und Angemessenheit der Aufgaben sichergestellt, dennoch handelt es sich nicht um standardisierte Verfahren. Auch die Durchführung der Tests als Gruppentestung kann als Schwäche gegenüber einer Individualtestung angesehen werden, da eine Gruppentestung nicht vollständig kontrollierbar ist, wodurch nicht für alle Kinder identische Rahmenbedingungen vorlagen. So variierten beispielsweise die Möglichkeiten des Blickkontakts zu anderen Schülern oder die Entfernung zur Tafel. Außerdem konnten die Schulstunde, in der getestet wurde oder die exakte Gruppengröße nicht über alle Schulen hinweg konstant gehalten werden, was ebenfalls eine Ergebniskonfundierung bedeuten kann. Diese Aspekte begründen zum Teil die niedrigeren Reliabilitäten der kognitiven Skalen im Vergleich zu den Motivationsfaktoren. Insgesamt lassen jedoch das Korrelationsmuster, sowie die prädiktive Validität der kognitiven Skalen den Schluss zu, dass nicht von einem substantiellen Messfehler auszugehen ist.

Hinsichtlich der Generalisierbarkeit der Befunde ist festzustellen, dass sich die Stichprobe bezüglich der Verteilung des höchsten elterlichen Bildungsabschlusses als weitestgehend repräsentativ erweist. Dennoch ist auch diese Stichprobe mit einer gewissen Selektivität behaftet, da nicht alle Viertklässler, die die zehn kooperierenden Grundschulen besuchten, auch an der Studie teilnahmen. Ca. 50% der Eltern entschieden sich gegen eine Teilnahme, so dass sich die Frage stellt, ob ein systematischer Ausfall vorlag. Über diesen Aspekt kann nur spekuliert werden, da keine diesbezüglichen Angaben zugänglich sind. Zwar äußerten die Lehrer den Verdacht, dass besonders Kinder aus höheren Schichten nicht an der Studie teilnahmen, da diese Eltern einer Datenerhebung in der Tendenz skeptisch gegenüberstanden. Allerdings spiegelt sich diese Vermutung nicht in einer Unterrepräsentation der oberen Bildungsschicht innerhalb der Stichprobe wider. Da

in vergleichbaren Studien außerdem häufig gerade niedrigere Sozialschichten unterrepräsentiert sind, ist es als ausgesprochen positiv zu bewerten, dass sich eine solche Verschiebung in der vorliegenden Studie weniger deutlich zeigt.

Doch es ist nicht nur die Selektion der Stichprobe als Teilmenge der Population zu erwähnen, denn auch innerhalb der getesteten Stichprobe fand eine Selektion statt. Nicht alle kognitiv getesteten Kinder beantworteten den Fragebogen, allerdings ist der Drop-out an dieser Stelle mit etwa 10% als relativ gering einzuschätzen. Dennoch kann nicht ausgeschlossen werden, dass es systematische Unterschiede zwischen den beiden Subgruppen gibt. Auch wenn sich keine signifikanten Unterschiede hinsichtlich der kognitiven Fähigkeiten der Kinder nachweisen lassen, ist zu vermuten, dass die eher motivierten und selbstsicheren Kinder bereit waren, den Fragebogen auszufüllen, was an einen gewissen zeitlichen Aufwand gebunden war. Dies gilt ebenso für die Elternstichprobe. Allerdings erweisen sich die Mittelwerte der motivationalen Skalen als mit anderen Studien vergleichbar, so dass diese Vermutung nicht belegt werden kann.

Generell gilt zu beachten, dass bei jeder Fragebogenuntersuchung, auch unter anonymisierten Bedingungen, die Problematik des sozial erwünschten Antwortverhaltens berücksichtigt werden sollte. So könnten sich die Kinder beispielsweise als besonders fähig und interessiert dargestellt haben, etwa aus der Vermutung heraus, dass Eltern und Lehrer die Antworten überprüfen. Dieses Antwortverhalten könnte so zu der Zustimmungstendenz der Kinder beigetragen haben. Dennoch erscheinen die Verteilungsmuster insgesamt als schlüssig, weshalb davon ausgegangen werden kann, dass dieser Aspekt vermutlich keine gravierende Ergebnisverzerrung verursacht hat. Interessant ist an dieser Stelle die Frage, ob das in älteren Stichproben nachgewiesene Absinken des FSKs bzw. der IM ebenfalls durch den Aspekt der sozialen Erwünschtheit mit verursacht wird. So ist denkbar, dass im pubertierenden Alter eine Verschiebung der Wertschätzung der Institution Schule im negativen Sinne einsetzt, wodurch es als sozial unerwünscht gelten könnte, eine positive Einstellung gegenüber der Schule zu äußern und Freude am Unterricht zu empfinden. Eine hohe Fähigkeitseinschätzung könnte ebenfalls sozial negativ bewertet werden, etwa wenn diese mit dem Konzept des Strebers in Verbindung gebracht wird.

Im Anschluss an die Darstellung der Grenzen der vorliegenden Studie werden nun verschiedene Punkte angesprochen, die in zukünftigen Studien berücksichtigt werden könnten.

5.9 Ausblick

Aus der Zusammenschau der Ergebnisdiskussion bzw. der Kritikpunkte der vorliegenden Studie, ergeben sich verschiedene Forschungsfragen, die in weiteren Studien Anwendung finden könnten.

Neben der Umsetzung eines längsschnittlichen Designs ist vor allem die Betrachtung der Zusammenhänge in einem genetisch sensitiven Design vielversprechend. Denn unter dem Gesichtspunkt, dass die Untersuchung des Einflusses des SES immer vor der Hintergrund der genetischen Kopplung von Intelligenz und SES der Eltern zu erfolgen hat, stellt die Methode der Verhaltensgenetik vermutlich einen vorteilhaften Untersuchungsansatz dar, verglichen mit einer querschnittlichen phänotypischen Analyse der Zusammenhänge. Allerdings ist zu bedenken, dass dieser Ansatz mit einer aufwendigen Stichprobengenerierung einhergeht.

Auf Konstruktebene ist zu sagen, dass die vergleichende Betrachtung von WM und gf als Prädiktoren schulischen Erfolgs weiterer Studien bedarf. Dabei scheint es zielführend, deutlicher auf Prozessebene differenzieren zu können, beispielsweise durch Einbezug von Aufgaben, die stärker auf spezifischen Aufmerksamkeitskontrollprozessen beruhen. Auch die Fokussierung auf spezifische exekutive Funktionen, wie Shifting, Inhibition oder Updating, spielt in der Forschung eine zunehmend wichtige Rolle, so dass auch diese Faktoren als potenzielle Variablen für weitere Analysen zu diskutieren wären. Insgesamt scheint die Forschung hier noch am Anfang zu stehen, so dass weitere Studien erforderlich sind, um die Beziehung zwischen WM, gf und Schulerfolg vollständig aufzuklären. Dabei scheint besonders der diagnostische Wert der WM-Erfassung für die schulische Praxis relevant zu sein (Carretti et al., 2009). Zu beachten ist hierbei, dass Defizite im Arbeitsgedächtnis häufig mit spezifischen Merkmalen wie Unaufmerksamkeit oder leichter Ablenkbarkeit einhergehen und teilweise auf spezifische WM-Komponenten (beispielsweise ZE und PL) zurückgeführt werden können, allerdings nicht automatisch auf mangelnder Intelligenz beruhen (Gathercole et al., 2008).

Außerdem gilt, dass die Betrachtung individueller Bedingungsfaktoren schulischer Leistung nicht erschöpfend ist (siehe Kapitel 1.5). Eine umfassende Erfassung individueller und familiärer Determinanten wäre eine wichtige Erweiterung. Denn die gleichzeitige Berücksichtigung einflussreicher Umweltvariablen, wie etwa Involvement oder Erziehungsverhalten (beispielsweise Seginer, 2006; Son & Morrison, 2010), ist vor dem Hintergrund des angestrebten umfassenden Verständnisses der komplexen Wirkzusammenhänge, die den Schulerfolg eines Schülers bedingen, erstrebenswert. Schließlich geben Studienergebnisse, die zeigen, dass sich elterliches Involvement positiv auf die wahrgenommene Autonomie und Leistungsmotivation eines Schülers auswirken kann, ebenso wie die Tatsache, dass ein hohes Maß an Kontrolle und Leistungsdruck zu negativen Effekten bezüglich der schulischen Leistung führen kann (Levpuscek & Zupancic, 2009), Anlass zur Annahme, dass elterliche Verhaltensweisen und die selbsteingeschätzten Fähigkeitsüberzeugungen und Wertvorstellungen der Kinder direkt verschaltet sind. So ist es schlüssig, diese Wirkzusammenhänge ebenfalls zu beleuchten.

Weiterhin bleibt festzustellen, dass der Vergleich der Studienergebnisse mit einer älteren Stichprobe wichtige Erkenntnisse liefern könnte. So stellt sich die Frage, welche Auswirkungen der Schulwechsel nach dem Ende der Grundschulzeit mit sich bringt. Die enormen Herausforderungen, die der Wechsel der Schule an die Schüler stellt, lässt die Frage aufkommen, ob sich die Gewichtung des Einflusses kognitiver und non-kognitiver Faktoren zu Gunsten der kognitiven Fähigkeiten verschiebt, da aufgrund der höheren zeitlichen Gesamtbelastung weniger leicht mangelnde Fähigkeit durch Fleiß kompensiert werden kann. Oder ob gerade in dieser Phase motivationale Eigenschaften an Einfluss gewinnen, weil diese es erleichtern, den neuen Herausforderungen standzuhalten und sich schnell an die neuen Gegebenheiten anzupassen. Hier sollten sich auch Effekte der Persönlichkeit zeigen, wie etwa der Ausprägung an Gewissenhaftigkeit. Während sich die Erfassung der Persönlichkeit im Grundschulalter als schwierig erweist, wird dieses Konstrukt in älteren Schülerstichproben häufig miterfasst.

Außerdem belegen zahlreiche Studien zum BFLPE (Marsh, 1987) (siehe Kapitel 1.3), dass durch die in den weiterführenden Schulen einsetzende Gruppierung nach Leistung verschiedene motivationale Prozesse ausgelöst werden, die mit dem veränderten Bezugsrahmen in Verbindung stehen und sich negativ auf das individuelle FSK auswirken können. Diese Effekte könnten jedoch durch die Einführung der Gemeinschaftsschulen abgemildert werden, was ebenfalls eine interessante Fragestellung für weitere Untersuchungen wäre. Allerdings erfordert die

Einbeziehung des Bezugsrahmens eine erheblich größere Stichprobengröße, da es sich dabei um ein genestetes Design handelt, in dem mindestens auf Schulebene, wenn nicht auf Klassenebene differenziert werden muss.

Die Untersuchung einer älteren Stichprobe hätte zusätzlich den Vorteil, dass weitere non-kognitive Faktoren, die in bisherigen Studien ihre Bedeutung unter Beweis gestellt haben (z.B. Zielorientierungen; Eccles & McGregor, 1999), valider zu erfassen wären als im Grundschulalter, was zu einem erweiterten Verständnis der Bedeutung motivationaler Facetten beitragen könnte. Hierzu könnten auch Variablen zählen, die stärker auf emotionale Aspekte fokussieren, wie beispielsweise Schulangst. Emotionen sind allgegenwärtig und treten daher auch im Schulsetting auf. Obwohl zahlreiche Motivationsmodelle emotionale Aspekte integrieren, ist auf empirischer Ebene eine Vernachlässigung dieser Facette festzustellen (Pekrun, 2009). Bedenkt man jedoch die Bandbreite an emotionalen Zuständen, die sich im Schulalltag manifestieren können, ist dies als nicht zufriedenstellend zu bezeichnen. So ist anzunehmen, dass sich maladaptive emotionale Zustände, wie gesteigerte Angst, Langeweile, Wut oder Hoffnungslosigkeit negativ auf Schulleistungen auswirken (Zeidner, 1998), während positive Zustände, wie Freude oder Stolz die Schulleistung steigern können (Pekrun, 2009). Zwar existieren bereits einige Studien bezüglich des Einflusses von Schulangst, aber insgesamt wäre eine Ausdehnung der Forschungsbestrebungen in Richtung emotionaler Einflussfaktoren wünschenswert. Schließlich ist zu beachten, dass Emotion, Kognition und Motivation konfundierte Konstrukte darstellen, die vielerlei Verknüpfungen aufweisen und sich somit wechselseitig beeinflussen (Pekrun, 2009).

Abschließend ist anzumerken, dass die nachgewiesenen Zusammenhänge stets vor dem Hintergrund der jeweiligen Kultur, aus der die Daten stammen, zu bewerten sind. So verweist auch das Erwartungs-Wert-Modell auf die Wirkung der kulturellen Umwelt auf die Erwartungs- und Wert-Komponenten (Eccles & Wigfield, 2002). Hieraus ergibt sich die Frage, ob sich das Verhältnis der relativen Bedeutung kognitiver und non-kognitiver Prädiktoren in anderen kulturellen Settings replizieren lässt, oder ob die Wirkmechanismen spezifischer Natur sind. Erste Hinweise auf kulturspezifische Unterschiede konnten in Kooperation mit der Chinese Academy of Sciences, Beijing bereits nachgewiesen werden (Lu, Weber, Spinath & Shi, 2011). Die Weiterverfolgung der vergleichenden Betrachtung zwischen den Kulturen ist sicherlich lohnenswert.

Zuletzt soll kurz auf die Frage eingegangen werden, welche praktischen Implikationen der Nachweis der prädiktiven Validität motivationaler Faktoren mit sich bringen kann.

5.10 Praktische Implikationen der Relevanz non-kognitiver Bedingungsfaktoren

Die Tatsache, dass bessere Schulnoten mit höherem FSK einhergehen, könnte so interpretiert werden, dass es sinnvoll ist, ein möglichst hohes, wenn auch unrealistisches FSK zu vermitteln. Die Vermittlung bzw. Aufrechterhaltung eines überhöhten FSKs ist jedoch kritisch zu bewerten, da dieses zur Folge haben kann, dass sich nicht mehr mit dem Lernstoff auseinandergesetzt wird, da die Überzeugung vorherrscht, die notwendigen Fähigkeiten bereits zu besitzen und Lernerfolg als selbstverständlich erachtet werden (Helmke, 1998). Vielmehr scheinen besonders solche Kinder erfolgreich zu sein, die ein mäßig überschätztes FSK aufweisen, vermutlich weil sie sich auch Aufgaben zutrauen und weiterverfolgen, die bei vollständig realistischer Einschätzung nicht angegangen würden (Helmke, 1998). Spinath (2005) fasst die diesbezüglichen Forschungsergebnisse derart zusammen, dass motivationale Kompetenzen besonders dann gefördert werden, wenn realistische Selbsteinschätzungen bestärkt werden, unrealistische Überzeugungen hingegen nicht verstärkt werden, wenn realistisch über Lernbedarf und Stärken informiert wird und die Folgen realistischer und unrealistischer Selbsteinschätzungen erfahrbar werden. Hintergrund der Bedeutung realistischer Einschätzungen ist, dass eine gute Passung von Fähigkeit und Aufgabenschwierigkeit Flow-Erleben, als Zustand optimaler Motivation, wahrscheinlicher macht. Eine realistische Selbsteinschätzung erleichtert nun die Wahl von angemessenen Aufgaben, welche gute Bedingungen für das Erleben von Stolz bereitstellen (z.B. Feather, 1967). Daraus schlussfolgert Spinath (2005), dass realistisches Feedback, das auch auf Schwächen hinweist, auf den ersten Blick zwar unangenehmer ist, als unrealistisch positives Feedback. Langfristig sollte es jedoch „aufgabenimmanent Anlass für positive Selbstbewertungen“ (Spinath, 2005, S. 210) bieten und damit zu einer positiven Bewertung angemessener Aufgaben führen. Übertragen auf den praktischen Alltag der Lehrer rät Spinath (2005) daher zu aufgabenbezogenen Rückmeldungen. Dieses Vorgehen soll verhindern, dass die Bewertung auf die

Person bezogen wird. Weiterhin gilt, dass unangemessen ausgewählte Aufgaben und unrealistische Selbsteinschätzungen nicht bestärkt werden sollten, was beispielsweise der Fall wäre, wenn Kinder für die Bewältigung zu leichter Aufgaben zu stark gelobt werden. Dieses Vorgehen erschwert die Entwicklung von Kompetenzen.

Die Fähigkeit zur realistischen Fähigkeitseinschätzung lässt sich besonders durch die Anwendung von individuellen, sachbezogenen Vergleichsmaßstäben erzielen (Spinath, 2005). Die Umsetzung dieser Leitlinien erfordert jedoch die diesbezügliche Schulung der Lehrkräfte. Außerdem bleibt offen, inwieweit sich aufgabenbezogene Rückmeldungen in den alltäglichen Schulablauf integrieren lassen, denn eine vollständige Abschaffung der Bewertung durch Noten ist als unrealistisch zu bezeichnen. Dennoch ist die Einbeziehung dieser aktuellen Erkenntnisse in den schulischen Alltag äußerst wünschenswert.

Literaturverzeichnis

- Ackerman, B. P., Brown, E. D. & Izard, C. E. (2004). The relations between contextual risk, earned income, and the school adjustment of children from economically disadvantaged families. *Developmental Psychology*, 40(2), 204–216.
- Ackerman, P. L. & Wolman, S. (2007). How accurate are self-estimates of intelligence? *Journal of Experimental Psychology: Applied*, 13, 57–78.
- Ackerman, P. L., Beier, M. E. & Boyle, M. O. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General*, 131(4), 567–589.
- Ackerman, P. L., Beier, M. E. & Boyle, M. O. (2005). Working Memory and Intelligence: the same or different constructs? *Psychological Bulletin*, 131(1), 30–60.
- Ahnert, J., Schneider, W. & Börs, K. (2008). Entwicklung motorischer Fähigkeiten vom Vorschul- bis ins frühe Erwachsenenalter. In W. Schneider (Hrsg.), *Entwicklung von der Kindheit bis zum Erwachsenenalter. Befunde der Münchner Längsschnittstudie LOGIK*. (S. 23–42). Weinheim: Beltz.
- Akaike, H. (1974). A new look at statistical model identification. *IEEE transactions on Automatic Control*, 19, 716–723.
- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52, 317–332.
- Alloway, T. P. (2009). Working Memory, but not IQ, predicts subsequent learning in children with learning difficulties. *European Journal of Psychological Assessment*, 25(2), 92–98.
- Alloway, T. P., Gathercole, S. E. & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: are they separable? *Child Development*, 77(6), 1698–1716.
- Alloway, T. P., Gathercole, S. E., Adams, A. M., Willis, C., Eaglen, R. & Lamont, E. (2005). Working memory and phonological awareness as predictors of progress towards early learning goals at school entry. *British Journal of Developmental Psychology*, 23, 417–426.
- Alloway, T. P., Gathercole, S. E., Willis, C. & Adams, A. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology*, 87, 85–106.
- Amado, A. (2003). Partitioning predicted variance into constituent parts: A primer on regression communality analysis. *Research in the schools*, 10(1), 91–97.

- Ames, C. & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology*, 80, 260-267.
- Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A. (2001). *Intelligenz-Struktur-Test 2000 R* [Intelligence-Structure-Test 2000 R]. Göttingen: Hogrefe.
- Arnett, P. A., Higginson, C. I., Voss, W. D., Bender, W. I., Wurst, J. M. & Tippin, J. M. (1999). Depression in multiple sclerosis: Relationship to working memory capacity. *Neuropsychology*, 13(4), 546-556.
- Artelt, C., Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Schümer, G., Stanat, P., Tillmann, K.-J. & Weiß, M. (Hrsg.). (2001). *PISA 2000: Zusammenfassung zentraler Befunde*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review*, 64, 359-372.
- Atkinson, R. C. & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & T. Janet (Hrsg.), *The psychology of learning and motivation*. Oxford, England: Academic Press.
- Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.
- Baddeley, A. D. (1998). The central executive: A concept and some misconceptions. *Journal of the International Neuropsychological Society*, 4, 523-526.
- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4, 417-423.
- Baddeley, A. D. (2006). Working memory: an overview. In S. Pickering (Hrsg.), *Working Memory and Education* (S. 1-31). New York: Academic Press.
- Baddeley, A. D. & Hitch, G. (1974). Working memory. In G. Bower (Hrsg.), *The psychology of learning and motivation* (S. 47-90). New York: Academic.
- Baddeley, A. D. & Logie, R. (1999). Working memory: The multiple component model. In A. Miyake & P. Shah (Hrsg.), *Models of working memory* (S. 28-61). New York: Cambridge University Press.
- Baddeley, A. D., Emslie, H., Kolodny, J. & Duncan, J. (1998). Random generation and the executive control of working memory. *Quarterly Journal of Experimental Psychology*, 51A, 819-852.
- Bartels, M., Rietveld, M. J. H., Van Baal, G. C. M. & Boomsma, D. I. (2002). Genetic and environmental influences on the development of intelligence. *Behavior Genetics*, 32, 237-249.

- Batty, G. D., Deary, I. J. & Gottfredson, L. S. (2007). Pre-morbid (early life) IQ and later mortality risk: Systematic review. *Annals of Epidemiology*, 17, 278-288.
- Baumert, J. (1995). Gender, science interest, teaching strategies and socially shared beliefs about gender roles in 7th graders – a multi-level analysis. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- Baumert, J. & Köller, O. (1998). Interest research in secondary level I: An overview. In L. Hoffmann, A. Krapp, K. A. Renninger & J. Baumert (Hrsg.), *Interest and learning* (S. 241–256). Kiel: IPN.
- Baumert, J. & Maaz, K. (2006). Das theoretische und methodische Rahmenkonzept von PISA zur Erfassung sozialer und kultureller Ressourcen der Herkunftsfamilie: Internationale und nationale Rahmenkonzeption. In J. Baumert, P. Stanat & R. Watermann (Hrsg.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit. Vertiefende Analysen im Rahmen von PISA 2000* (S. 11- 29). VS Verlag für Sozialwissenschaften.
- Baumert, J. & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 323-410). Opladen: Leske + Budrich.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J. & Weiß, M. (Hrsg.). (2001). *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Baumert, J., Watermann, R. & Schümer, G. (2003). Disparitäten der Bildungsbeteiligung und des Kompetenzerwerbs. Ein institutionelles und individuelles Mediationsmodell. *Zeitschrift für Erziehungswissenschaft*, 6, 46-72.
- Bayliss, D. M., Jarrold, C., Baddeley, A. D., Gunn, D. M. & Leigh, E. (2005). Mapping the developmental constraints on working memory span performance. *Developmental Psychology*, 41, 579–597.
- Bayliss, D. M., Jarrold, C., Gunn, D. M. & Baddeley, A. D. (2003). The Complexities of Complex Span: Explaining Individual Differences in Working Memory in Children and Adults. *Journal of Experimental Psychology: General*, 132(1), 71-92.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238–246.

- Berch, D. B. (2008). Working memory and mathematical cognitive development: Limitations of limited-capacity resource models. *Developmental neuropsychology*, 33(3), 427-446.
- Berti, S. (2010). Arbeitsgedächtnis: Vergangenheit, Gegenwart und Zukunft eines theoretischen Konstruktes. *Psychologische Rundschau*, 61(1), 3-9.
- Bickley, P. G., Keith, T. Z. & Wolfle, L. M. (1995). The Three-Stratum Theory of cognitive abilities: Test of the structure of intelligence across the life span. *Intelligence*, 20, 309–328.
- Binet, A. & Simon, T. (1905a). Applications des méthodes nouvelles au diagnostic du niveau intellectuel chez des enfants normaux et anormaux d'hospice et d'école primaire. *Année Psychologique*, 11, 245-336.
- Binet, A. & Simon, T. (1905b). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *Année Psychologique*, 11, 411-463.
- Binet, A. & Simon, T. (1905c). Sur la nécessité d'établir un diagnostic scientifique des états inférieurs de l'intelligence. *Année Psychologique*, 11, 163-190.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bornstein, M. H. & Bradley, R. H. (2003). *Socioeconomic Status, Parenting, and Child Development*. Mahwah: Erlbaum.
- Bornstein, M. H., Hahn, C.- S., Suwalsky, J. T. D. & Haynes, O. M. (2003). Socioeconomic Status, Parenting, and Child Development: The Hollingshead Four- Factor- Index of Occupations. In M. H. Bornstein & R. H. Bradley (Hrsg.), *Socioeconomic Status, Parenting, and Child Development* (S. 29- 82). Mahwah: Erlbaum.
- Bos, W., Hornberg, S., Arnold, K.-H., Faust, G., Fried, L., Lankes, E.-M., Schwippert, K. & Valtin, R. (Hrsg.). (2008). *IGLU-E 2006. Die Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich*. Münster: Waxmann.
- Bouchard, T. J. & McGue, M. (2003). Genetic and environmental influences on human psychological differences. *Journal of Neurobiology*, 54, 4–45.
- Boulanger, F. D. (1981). Ability and science learning: A quantitative synthesis. *Journal of Research in Science Teaching*, 18, 113-121.
- Bradley, R. H. & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, 53, 371-399.
- Bradley, R. H., Corwyn, R. F., Burchinal, M., McAadoo, H .P. & Garcia Coll, C. (2001). The home environments of children in the United States. Part 2: Relations with behavioral development through age 13. *Child Development*, 72, 1868-1886.

- Brocke, B. & Beauducel, A. (2001). Intelligenz als Konstrukt. In E. Stern & J. Guthke (Hrsg.), *Perspektiven der Intelligenzforschung* (S. 13-42). Lengerich: Pabst Science Publishers.
- Brody, N. (1992). *Intelligence*. New York: Academic Press.
- Brody, G. H. & Flor, D. L. (1998). Maternal resources, parenting practices, and child competence in rural, single-parent African American families. *Child Development*, 69(3), 803–816.
- Brooks-Gunn, J. & Duncan, G. J. (1997). The Effects of Poverty on Children. *The Future of Children*, 7 (2), Children and Poverty (Summer- Autumn), 55-71.
- Browne, M. W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Hrsg.), *Testing structural equation models* (S. 136–162). Newbury Park, CA: Sage.
- Brunner, M. (2006). *Mathematische Schülerleistung: Struktur, Schulformunterschiede und Validität*. Berlin: Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II.
- Bundesministerium für Bildung und Forschung (2007). *Rahmenprogramm zur Förderung der empirischen Bildungsforschung*. Abgerufen am 02.01.2012 unter http://www.bmbf.de/pubRD/foerderung_der_empirischen_bildungsforschung.pdf.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson.
- Bull, R. & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology*, 19(3), 273-293.
- Bull, R., Espy, K. A. & Wiebe, S. A. (2008). Short-term memory, Working Memory, and Executive Functioning in Preschoolers: Longitudinal Predictors of Mathematical Achievement at Age 7 Years. *Developmental Neuropsychology*, 33(3), 205-228.
- Bull, R., Johnston, R. S. & Roy, J. A. (1999). Exploring the roles of the visual-spatial sketchpad and central executive in children's arithmetical skills: Views from cognition and developmental neuropsychology. *Developmental Neuropsychology*, 15, 421–442.
- Byrne, M. B. (2001). *Structural equation modeling with AMOS. Basic concepts, applications, and programming*. London: Lawrence Erlbaum.
- Calsyn, R. & Kenny, D. (1977). Self-concept of ability and perceived evaluations by others: Cause or effect of academic achievement? *Journal of Educational Psychology*, 69, 136-145.

- Campbell, T., Dollaghan, C., Needleman, H. & Janosky, J. (1997). Reducing bias in language assessment: Processing dependent measures. *Journal of Speech, Language, and Hearing Research*, 40(3), 519–525.
- Carretti, B., Borella, E., Cornoldi, C. & De Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learning and Individual Differences*, 19, 246-251.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- Case, R., Kurland, D. M. & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, 33, 386-404.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54(1), 1-22.
- Cattell, R. B. (1973). *Measuring intelligence with the Culture Fair tests*. Champaign, IL: Institute for Personality and Ability Testing.
- Chamorro-Premuzic, T. (2007). *Personality and individual differences*. Oxford: Blackwell.
- Chamorro-Premuzic, T., Harlaar, N., Greven, C. U. & Plomin, R. (2010). More than just IQ: A longitudinal examination of self-perceived abilities as predictors of academic performance in a large sample of UK twins. *Intelligence*, 38, 385-392.
- Chen, F., Bollen, K., Paxton, P., Curran, P. J. & Kirby, J. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods and Research*, 29, 468–508.
- Clark, C. A. C., Pritchard, V. E. & Woodward, L. J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Developmental Psychology*, 46(5), 1176-1191.
- Colom, R. & Flores-Mendoza, C. E. (2007). Intelligence predicts scholastic achievement irrespective of SES factors: Evidence from Brazil. *Intelligence*, 35, 243– 251.
- Colom, R., Flores-Mendoza, C. & Rebollo, I. (2003). Working memory and intelligence. *Personality and Individual Differences*, 34(1), 33–39.
- Colom, R., Abad, F. J., Quiroga, A., Shih, P. C. & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs, but why? *Intelligence*, 36, 584-606.
- Conway, A. R. A., Kane, M. J. & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, 7(12), 547-552.

- Conway, A. R. A., Cowan, N., Bunting, M. F., Theriault, D. J. & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, 30, 163-183.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O. & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5), 769-786.
- Cooley, W. W. & Lohnes, P. R. (1976). *Evaluation research in education*. New York: Irvington.
- Covington, M. V. (1992). *Making the grade: A self-worth perspective in motivation and school reform*. New York, NY, US: Cambridge University Press.
- Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin*, 104, 163-191.
- Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Hrsg.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (S. 62–101). Cambridge, UK: Cambridge University Press.
- Cowan, N., Day, L., Saults, J. S., Keller, T. A., Johnson, T. & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language*, 31, 1–17.
- Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., et al. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, 51(1), 42–100.
- Cowan, N., Wood, N. L., Wood, P. K., Keller, T. A., Nugent, L. D. & Keller, C. V. (1998). Two separate verbal processing speeds contributing to verbal short-term memory span. *Journal of Experimental Psychology: General*, 127, 141–160.
- Csikszentmihalyi, M. (1988). The flow experience and its significance for human psychology. In M. Csikszentmihalyi & I. S. Csikszentmihalyi (Hrsg.), *Optimal experience: Psychological studies of flow in consciousness* (S. 15-35). New York, US: Cambridge University Press.
- Curran, P. J., West, S. G. & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1, 16–29.
- Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466.

- Daneman, M. & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Verbal Learning and Verbal Behavior*, 25, 1–18.
- deCharms, R. (1968). *Personal Causation: The Internal Affective Determinants of Behavior*. New York: Academic.
- De Jong, P. P. (1998). Working memory deficits of reading disabled children. *Journal of Experimental Child Psychology*, 70(2), 75–96.
- De Jong, P. & de Jong, P. F. (1996). Working memory, intelligence and reading ability in children. *Personality and Individual Differences*, 21(6), 1007-1020.
- De Jong, P. F. & Das-Smaal, E. A. (1995). Attention and Intelligence: The Validity of the Star Counting Test. *Journal of Educational Psychology*, 87(1), 80-92.
- De Smedt, B., Janssen, R., Bouwens, K., Verschaffel, L., Boets, B. & Ghesquière, P. (2009). Working memory and individual differences in mathematics achievement: A longitudinal study from first grade to second grade. *Journal of Experimental Child Psychology*, 103, 186-201.
- Deary, I. J. (2001). *Intelligence. A very short introduction*. Oxford: Oxford University Press.
- Deary, I. J., Penke, L. & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews Neuroscience*, 11, 201-211.
- Deary, I. J., Strand, S., Smith, P. & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35, 13-21.
- Deary, I. J., Whalley, L. J., Lemmon, H., Crawford, J. R. & Starr, J. M. (2000). The stability of individual differences in mental ability from childhood to old age. *Intelligence*, 28, 49–55.
- Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J. & Fox, H. C. (2004). The impact of childhood intelligence in later life: Following up the Scottish Mental Surveys of 1932 and 1947. *Journal of Personality and Social Psychology*, 86, 130–147.
- Deci, E. L. & Ryan, R. M. (1985). *Intrinsic Motivation and self-determination in human behaviour*. New York: Plenum.
- Deci, E. L. & Ryan, R. M. (2008). Self-Determination Theory: A Macrotheory of Human Motivation, Development, and Health. *Canadian Psychology*, 49(3), 182-185.
- Della Sala, S., Gray, C., Baddeley, A., Allamano, N. & Wilson, L. (1999). Pattern span: a tool for unwinding visuo-spatial memory. *Neuropsychologia*, 37, 1189–1199.
- Demmer, M. (2009). Arbeiterkinder müssen Superschüler sein... um nach der Grundschule aufs Gymnasium zu gehen. *Erziehung und Wissenschaft*, 61(1), 18-19.

- Dempster, F. M. (1985). Short-term memory development in childhood and adolescence. In C. J. Brainerd & M. Pressley (Hrsg.), *Basic processes in memory development* (S. 209–248). New York: Springer.
- Ditton, H. (1992). *Ungleichheit und Mobilität durch Bildung. Theorie und empirische Untersuchung über sozialräumliche Aspekte von Bildungsentscheidungen*. Weinheim und München: Juventa.
- Duff, S. C. & Logie, R. H. (2001). Processing and storage in working memory span. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 54A, 31–48.
- Duncan, G. J. & Magnuson, K. A. (2003). Off with Hollingshead: Socioeconomic Resources, Parenting, and Child Development. In M. H. Bornstein & R. H. Bradley (Hrsg.), *Socioeconomic Status, Parenting, and Child Development* (S. 83–106). Mahwah: Erlbaum.
- Duncan, G. J., Brooks-Gunn, J. & Klebanov, P. K. (1994). Economic Deprivation and Early Childhood Development. *Child Development*, 65, 296–318.
- Dupeyrat, C. & Marine, C. (2005). Implicit theories of intelligence, goal orientations, cognitive engagement, and achievement: A test of Dwecks' model with returning to school adults. *Contemporary Educational Psychology*, 30, 43–59.
- Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, 41, 1040–1048.
- Dweck, C. S. & Elliott, E. S. (1983). Achievement motivation. In P. H. Mussen & E. M. Hetherington (Hrsg.), *Handbook of child psychology: Vol. IV. Social and personality development* (S. 643–691). New York: Wiley.
- Dweck, C. S. & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95(2), 256–273.
- Dweck, C. S. & Master, A. (2009). Self-theories and motivation: student's beliefs about intelligence. In K. R. Wentzel & A. Wigfield (Hrsg.), *Handbook of motivation at school* (S. 123–144). New York: Routledge.
- Eccles, J. S. (1984). Sex Differences in Achievement Patterns. In Sonderegger, T. B. (Hrsg.), *Nebraska Symposium on Motivation Volume 32* (S. 97–132). Lincoln: University of Nebraska Press.
- Eccles, J. S. (2005). Subjective Task Value and the Eccles et al. Model of Achievement-Related Choices. In A. J. Elliot & C. S. Dweck (Hrsg.), *Handbook of Competence and Motivation* (S. 105–121). New York: The Guilford Press.

- Eccles, J. S. & Midgley, C. (1989). Stage/environment fit: Developmentally appropriate classrooms for early adolescents. In R. E. Ames & C. Ames (Eds.), *Research on motivation in education* (Vol. 3). New York: Academic Press.
- Eccles, J. S. & Wigfield, A. (1995). In the Mind of the Actor: The Structure of Adolescents' Achievement Task Values and Expectancy-Related Beliefs. *Personality and Social Psychology Bulletin*, 21, 215-225.
- Eccles, J. S. & Wigfield, A. (2002). Motivational Beliefs, Values, and Goals. *Annual Review of Psychology*, 53, 109-132.
- Eccles, J. S., Adler, T. & Meece, J. L. (1984). Sex Differences in Achievement: A Test of Alternate Theories. *Journal of Personality and Social Psychology*, 46, 26-43.
- Eccles, J., Midgley, C. & Adler, T. (1984). Grade-related changes in the school environment: Effects on achievement motivation. In J. G. Nicholls (Hrsg.), *The development of achievement motivation* (S. 283-331). Greenwich, CT: JAI Press.
- Eccles, J. S., O'Neill, S. A. & Wigfield, A. (2005). Ability self-perceptions and subject task values in adolescents and children. In K. A. Moore & L. H. Lippman (Hrsg.), *What do children need to flourish? Conceptualizing and measuring indicators of positive development* (S. 237-249). New York, NY: Springer.
- Eccles, J. S., Wigfield, A. & Blumenfeld, P. C. (1984). *Psychological predictors of competence development* (Grant No. 2 ROI HD 17553-01). Bethesda, MD: National Institute of Child Health and Human Development.
- Eccles, J. S., Wigfield, A., Harold, R. D. & Blumenfeld, P. (1993). Age and Gender Differences in Children's Self- and Task Perceptions during Elementary School. *Child Development*, 64, 830-847.
- Eccles (Parsons), J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L. & Midgley, C. (1983). Expectancies, Values, and Academic Behaviors. In J. T. Spence (Hrsg.), *Achievement and Achievement Motives* (S. 75-146). San Francisco: Freeman.
- Eccles, J. S., Wigfield, A. & Schiefele, U. (1998). Motivation to succeed. In W. Damon & N. Eisenberg (Hrsg.), *Handbook of child psychology, 6th ed, vol 3, social, emotional and personality development* (S. 1017-1095). New York: Wiley.
- Eder, F. (1992). Schulklima und Entwicklung allgemeiner Interessen. In A. Krapp & M. Prenzel (Hrsg.), *Interesse, Lernen, Leistung* (S. 165-194). Münster: Aschendorff.
- Ehmke, T., Siegle, T. & Hohensee, F. (2005). Soziale Herkunft im Ländervergleich. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der zweite Vergleich der Länder in Deutschland – Was wissen und können Jugendliche?* (S. 235-268). Münster: Waxmann.

- Ehmke, T., Hohensee, F., Heidemeier, H. & Prenzel, M. (2004). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 225-254). Münster: Waxmann.
- Elliott, A. J. & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of personality and social psychology*, 72, 218-232.
- Elliot, A. J. & Dweck, C. S. (Eds.). (2005). *Handbook of competence and motivation*. New York: Guilford.
- Elliot, A. J. & McGregor, H. A. (1999). Test anxiety and the hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology*, 76, 628–644.
- Elliot, A. J. & McGregor, H. A. (2001). A 2x2 Achievement Goal Framework. *Journal of Personality and Social Psychology*, 80(3), 501-519.
- Elliot, A. J., McGregor, H. A. & Gable, S. (1999). Achievement goals, study strategies, and exam performance: A mediational analysis. *Journal of Educational Psychology*, 91, 549–563.
- EllisWeismer, S., Tomblin, J. B., Zhang, X. Y., Buckwalter, P., Chynoweth, J. G. & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research*, 43(4), 865–878.
- Engel de Abreu, P. M. J., Conway, A. R. A. & Gathercole, S. E. (2010). Working memory and fluid intelligence in young children. *Intelligence*, 38, 552-561.
- Engel, P. M. J., Santos, F. H. & Gathercole, S. E. (2008). Are working memory measure free of socioeconomic influence? *Journal of Speech, Language, and Hearing Research*, 51, 1580-1587.
- Engle, R. W., Kane, M. J. & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex. In A. Miyake & P. Shah (Hrsg.), *Models of working memory* (S. 102-134). Cambridge, England: Cambridge University Press.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E. & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, 128, 309–331.

- Ensminger, M. E. & Fothergill, K. E. (2003). A Decade of Measuring SES: What it tells us and where to go from here. In M. H. Bornstein & R. H. Bradley (Hrsg.), *Socioeconomic Status, Parenting, and Child Development* (S. 13- 27). Mahwah: Erlbaum.
- Espy, K. A., McDiarmid, M. M., Cwik, M. F., Stalets, M. M., Hamby, A. & Senn, T. E. (2004). The contribution of executive functions to emergent mathematic skills in preschool children. *Developmental Neuropsychology*, 26, 465–486.
- Evans, G. W., Gonnella, C., Marcynyszyn, L. G & Salpekar, N. (2005). The role of chaos in poverty and children's socioemotional adjustment. *Psychological Science*, 16, 560–565.
- Farah, M. J., Hammond, K. M., Levine, D. N. & Calvanio, R. (1988). Visual and spatial mental imagery: Dissociable systems of representation. *Cognitive Psychology*, 20(4), 439-462.
- Feather, N. T. (1967). Valence of outcome and expectations of success in relation to task difficulty and perceived locus of control. *Journal of personality and social psychology*, 7, 372-386.
- Fleming, M. L. & Malone, M. R. (1983). The relationship of student characteristics and student performance in science as viewed by meta-analysis research. *Journal of Research in Science Teaching*, 20, 481-495.
- Fraser, B. J., Walberg, H. J., Welch, W. W. & Hattie, J. A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research*, 11(2), 147-252.
- Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, 17, 172–179.
- Fry, A. & Hale, S. (2000). Relationship among processing speed, working memory, and fluid intelligence in children. *Biological Psychology*, 54, 1-34.
- Fuligni, A. J. (1997). The academic achievement of adolescents from immigrant families: The roles of family background, attitudes, and behavior. *Child Development*, 68(2), 351–363.
- Gagné, F. & St Père, F. (2001). When IQ is controlled, does motivation still predict achievement? *Intelligence*, 30, 71-100.
- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

- Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, 3, 410–418.
- Gathercole, S. E. (2001), Ed.. *Short-term and Working Memory*. Psychology Press. Hove: England.
- Gathercole, S. E. & Hitch, G. J. (1993). Developmental changes in short-term memory: A revised working memory perspective. In A. Collins, S. E. Gathercole, M. A. Conway & P. E. Morris (Hrsg.), *Theories of memory* (S. 189–210). Hove, England: Erlbaum.
- Gathercole, S. E. & Pickering, S. J. (2000a). Assessment of Working Memory in Six- and Seven-Year-Old Children. *Journal of Educational Psychology*, 92(2), 377–390.
- Gathercole, S. E. & Pickering, S. J. (2000b). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *British Journal of Educational Psychology*, 70, 177–194.
- Gathercole, S. E., Brown, L. & Pickering, S. J. (2003). Working memory assessments at school entry as longitudinal predictors of National Curriculum attainment levels. *Educational and Child Psychology*, 20, 109–122.
- Gathercole, S.E., Pickering, S.J., Ambridge, B. & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, 40(2), 177–190.
- Gathercole, S.E., Pickering, S.J., Knight, C. & Stegmann, Z. (2004). Working Memory Skills and Educational Attainment: Evidence from National Curriculum Assessments at 7 and 14 Years of Age. *Applied Cognitive Psychology*, 18, 1–16.
- Gathercole, S. E., Alloway, T. P., Kirkwood, H. J., Elliott, J. G., Holmes, J. & Hilton, K. A. (2008). Attentional and executive function behaviours in children with poor working memory. *Learning and Individual Differences*, 18, 214–223.
- Geary, D. C., Hamson, C. O. & Hoard, M. K. (2000). Numerical and arithmetical cognition: A longitudinal study of process and concept deficits in children with learning disability. *Journal of Experimental Child Psychology*, 77, 236–263.
- Giedd, J. N., Schmitt, J. E. & Neale, M. C. (2007). Structural brain magnetic imaging of pediatric twins. *Human Brain Mapping*, 28, 474–481.
- Giesen, H., Gold, A., Hummer, A. & Jansen, R. (1986). *Prognose des Studienerfolgs. Ergebnisse aus Längsschnittuntersuchungen*. Frankfurt: Universität Frankfurt, Institut für Pädagogische Psychologie.
- Gomes, H., Molholm, Christodoulou, C., Ritter, W. & Cowan, N. (2000). The development of auditory attention in children. *Frontiers in Bioscience* 2000, 5, 108–120.

- Gottfredson, L. S. (1997a). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, 24, 13-23.
- Gottfredson, L. S. (1997b). Why g matters: the complexity of everyday life. *Intelligence*, 24, 79–132.
- Gottfredson, L. S. (1998). The general intelligence factor. *Scientific American presents*, 9(4), 24-29.
- Gottfredson, L. S. (2002). Where and why g matters: Not a mystery. *Human performance*, 15, 25-46.
- Gottfredson, L. S. & Saklofske, D. H. (2009). Intelligence: Foundations and Issues in Assessment. *Canadian Psychology*, 50(3), 183-195.
- Gottfried, A. E. (1985). Academic intrinsic motivation in elementary and junior high school students. *Journal of Educational Psychology*, 77, 631–645.
- Gottfried, A. E. (1990). Academic intrinsic motivation in young elementary school children. *Journal of Educational Psychology*, 82(3), S. 525-538.
- Gottfried, A. W., Gottfried, A. E., Bathurst, K., Guerin, D. W. & Parramore, M. (2003). Socioeconomic status in children's development and family environment: Infancy through adolescence. In M. Bornstein & R. Bradley (Hrsg.), *Socioeconomic Status, Parenting, and Child Development*. Mahway, NJ: Lawrence Erlbaum.
- Greene, B. A. & Miller, R. B. (1996). Influences on achievement: Goals, perceived ability, and cognitive engagement. *Contemporary Educational Psychology*, 21, 181–192.
- Greven, C. U., Harlaar, N., Kovas, Y., Chamorro-Premuzic, T. & Plomin, R. (2009). More Than Just IQ. School Achievement Is Predicted by Self-Perceived Abilities – But for Genetic Rather Than Environmental Reasons. *Psychological Science*, 20(6), 753-762.
- Grube, D. & Hasselhorn, M. (2006). Längsschnittliche Analysen zur Lese-, Rechtschreib- und Mathematikleistung im Grundschulalter: zur Rolle von Vorwissen, Intelligenz, phonologischem Arbeitsgedächtnis und phonologischer Bewusstheit. In I. Hosenfeld & F.-W. Schrader (Hrsg.), *Schulische Leistung. Grundlagen, Bedingungen, Perspektiven* (S. 87-105). Münster: Waxmann.
- Gruber, O., Indefrey, P., Steinmetz, H. & Kleinschmidt, A. (2001). Dissociating neural correlates of cognitive components in mental calculation. *Cerebral Cortex*, 11, 350–359.
- Guay, F., Marsh, H. W. & Boivin, M. (2003). Academic Self-Concept and Academic Achievement: Developmental Perspectives on Their Causal Ordering. *Journal of Educational Psychology*, 95(1), 124-136.

- Guay, F., Ratelle, C. F. & Chanal, J. (2008). Optimal learning in optimal contexts: The role of self-determination in education. *Canadian Psychology*, 49, 233–240.
- Gustafsson, J. E. & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, 28, 407-434.
- Gustafsson, J. E., & Undheim, J. O. (1996). Individual differences in cognitive functions. In D. C. Berliner, & R. C. Calfee (Hrsg.), *Handbook of Educational Psychology* (S. 186–242). New York: Prentice Hall International.
- Hamburg-Wechsler-Intelligenztest für Kinder IV (HAWIK IV) (2008). F. Petermann & U. Petermann (Hrsg.). Bern: Huber.
- Hanley, J. R., Young, A. W. & Pearson, N. A. (1991). Impairment of the visuo-spatial sketch pad. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 43A(1), 101-125.
- Hany, E. A. (2001). Die Vererbung der Intelligenz unter der Entwicklungsperspektive. In E. Stern & J. Guthke (Hrsg.), *Perspektiven der Intelligenzforschung* (S. 69-88). Lengerich: Pabst.
- Hart, S. A., Petrill, S. A., Deater-Deckard, K. & Thompson, L. A. (2007). SES and CHAOS as environmental mediators of cognitive ability: A longitudinal genetic analysis. *Intelligence*, 35, S. 233–242.
- Hattie, J. A. & Hansford, B. C. (1982). *Personality and intelligence: What relationship with achievement?* Paper presented at the Annual Conference of Australian Association for Research in Education, Brisbane.
- Hebb, D. O. (1955). Drives and the C.N.S. (conceptual nervous system). *Psychological Review*, 62, 243–54.
- Hecht, S. A., Torgesen, J. K., Wagner, R. K. & Rashotte, C. A. (2001). The relations between phonological processing abilities and emerging individual differences in mathematical computation skills: A longitudinal study from second to fifth grades. *Journal of Experimental Child Psychology*, 79, 192–227.
- Heckhausen, H. (1984). Emergent achievement behavior: Some early developments. In J. Nicholls (Hrsg.), *Advances in achievement motivation* (S. 1-32). Greenwich, CT: JAI Press.
- Heckhausen, H. (1987). Emotional components of action: Their ontogeny as reflected in achievement behavior. In D. Giritz & J. F. Wohlwill (Hrsg.), *Curiosity, imagination, and play*. Hillsdale, NJ: Erlbaum.
- Heckhausen, H. (2006). Entwicklungslinien der Motivationsforschung. In J. Heckhausen & H. Heckhausen (Hrsg.), *Motivation und Handeln* (S. 11- 43), 3. Auflage. Berlin: Springer.

- Heckhausen J. & Heckhausen H. (2006a). Motivation und Handeln: Einführung und Überblick. In J. Heckhausen & H. Heckhausen (Hrsg.), *Motivation und Handeln* (S.1- 9), 3. Auflage. Berlin: Springer.
- Heller, K. A. (1997). Individuelle Bedingungsfaktoren der Schulleistung: Literaturüberblick. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 183-201). Weinheim: Beltz.
- Helmke, A. (1992). *Selbstvertrauen und schulische Leistungen*. Göttingen: Hogrefe.
- Helmke, A. (1993). Motivation und Schulleistung: Trends, Sackgassen und Perspektiven der Forschung. In L. Montada (Hrsg.), *Bericht über den 38. Kongreß der Deutschen Gesellschaft für Psychologie in Trier 1992* (Bd. 2, S. 588-595). München: Hogrefe.
- Helmke, A. (1998). Vom Optimisten zum Realisten? Zur Entwicklung des Fähigkeitsselbstkonzepts vom Kindergarten bis zur sechsten Klassenstufe. In F. E. Weinert (Hrsg.), *Entwicklung im Kindesalter* (S. 115-132). Weinheim: Beltz.
- Helmke, A. & Schrader, F.-W. (2006). Determinanten der Schulleistung. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 83-94). Weinheim: Beltz.
- Helmke, A. & van Aken, M. (1995). The causal ordering of academic achievement and selfconcept of ability during elementary school: A longitudinal study. *Journal of Educational Psychology*, 87, 624-637.
- Helmke, A. & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen. In Weinert, F. E. (Hrsg.), *Psychologie des Unterrichts und der Schule* (S. 71-176). Göttingen: Hogrefe.
- Hitch, G. J., Towse, J. N. & Hutton, U. (2001). What limits children's working memory span? Theoretical accounts and applications for scholastic development. *Journal of Experimental Psychology: General*, 130, 184–198.
- Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, 74(5), 1368–1378.
- Hoffmeyer-Zlotnik, J. H. P. (2003). „Stellung im Beruf“ als Ersatz für eine Berufsklassifikation zur Ermittlung von sozialem Prestige. *ZUMA- Nachrichten*, 27(52), 114- 127.
- Hofstätter, P. R. (1957). *Psychologie*. Frankfurt: Fischer.
- Hoge, R. D. & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, 59(3), 297-313.
- Holling, H., Preckel, F. & Vock, M. (2004). *Intelligenzdiagnostik*. Göttingen: Hogrefe.

- Holmes, J. & Adams, J. W. (2006). Working memory and children's mathematical skills: Implications for mathematical development and mathematics curricula. *Educational Psychology, 26*, 339–366.
- Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review, 75*, 242-259.
- Horn, J. L. & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology, 57*, 253-270.
- Horn, J. L. & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica, 26*, 107-129.
- Hornung, C., Brunner, M., Reuter, R. A. P. & Martin, R. (2011). Children's working memory: Its structure and relationship to fluid intelligence. *Intelligence, 39*, 210-221.
- Hulme, C., Thomson, N., Muir, C. & Lawrence, A. (1984). Speech rate and the development of short-term memory span. *Journal of Experimental Child Psychology, 38*, 241–253.
- Humphreys, L. G. & Stark, S. (2002). General intelligence: Measurement, correlates, and interpretations of the cultural-genetic construct. In R. J. Sternberg & E. Grigorenko (Hrsg.), *The general factor of intelligence. How general is it?* (S. 87-115). Mahwah: Erlbaum.
- Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M. & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27*(2), 236–248.
- Hutton, U. M. Z. & Towse, J. N. (2001). Short-term memory and working memory as indices of children's cognitive skills. *Memory, 9*(4–6), 383–394.
- Isaacs, E. B. & Vargha-Khadem, F. (1989). Differential course of development of spatial and verbal memory span: a normative study. *British Journal of Developmental Psychology, 7*, 377 - 380.
- Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S. & Wigfield, A. (2002). Changes in Children's Self-Competence and Values: Gender and Domain Differences across Grades One through Twelve. *Child Development, 73*(2), 509-527.
- Jäger, A. O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen. Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells. *Diagnostica, 28*, 195-226.

- Jäger, A. O., Holling, H., Preckel, F., Schulze, R., Vock, M., Süß, H.-M. & Beauducel, A. (2004). *Berliner Intelligenzstrukturtest für Jugendliche: Begabungs- und Hochbegabungsdiagnostik (BIS-HB)*. Göttingen: Hogrefe.
- Jarrold, C. & Towse, J. N. (2006). Individual Differences in Working Memory. *Neuroscience*, 139, 39-50.
- Jarvis, H. L. & Gathercole, S. E. (2003). Verbal and non-verbal working memory and achievements on national curriculum tests at 11 and 14 years of age. *Educational and Child Psychology*, 20(3), 123-140.
- Jencks, C. (1979). Who gets ahead? *The determinants of economic success in America*. New York: Basic Books.
- Jensen, A. R. (1993). Psychometric g and achievement. In B. R. Gifford (Hrsg.), *Policy perspectives on educational testing* (S. 117-227). Boston: Kluwer Academic.
- Jensen, A. R. (1998). *The g Factor: The Science of Mental Ability*. Westport: Praeger.
- Jeynes, W. H. (2002a). *Divorce, Family Structure and the Academic success of children*. The Haworth Press.
- Jeynes, W. H. (2002b). The Challenge of Controlling for SES in Social Science and Education Research. *Educational Psychology Review*, 14(2), 205- 221.
- Johnson, W., Corley, J., Starr, J. M. & Deary, I. J. (2011). Psychological and Physical Health at Age 70 in the Lothian Birth Cohort 1936: Links With Early Life IQ, SES, and Current Cognitive Function and Neighborhood Environment. *Health Psychology*, 30, 1–11.
- Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Hrsg.), *Testing structural equation models* (S. 294–316). Newbury Park, CA: Sage.
- Jöreskog, K. G. & Sörbom, D. (1989). *LISREL 7 user's reference guide*. Chicago: SPSS Publications.
- Joussemet, M., Landry, R. & Koestner, R. (2008). A self-determination theory perspective on parenting. *Canadian Psychology*, 49, 194–200.
- Jussim, L., Robustelli, S. L. & Cain, T. R. (2009). Teacher expectations and self-fulfilling prophecies. In K. R. Wentzel & A. Wigfield (Hrsg.), *Handbook of motivation at school* (S. 349-380). New York: Routledge.
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W. & Engle, R. W. (2004). The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. *Journal of Experimental Psychology: General*, 133(2), 189-217.

- Kemp, L. C. D. (1955). Environmental and other characteristics determining attainments in primary schools. *British Journal of Educational Psychology*, 25, 66–77.
- Kintsch, W., Healy, A. F., Hegarty, M., Pennington, B. F. & Salthouse, T. A. (1999). Models of working memory. Eight questions and some general issues. In A. Miyake & P. Shah (Hrsg.), *Models of working memory. Mechanisms of active maintenance and executive control* (S. 412-441). Cambridge: University Press.
- Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling*. New York: The Guilford Press.
- Köller, O., Baumert, J. & Schnabel, K. U. (2001). Does interest matter? The relationship between academic interest and achievement in mathematics. *Journal of Research in Mathematics Education*, 32, 448–470.
- Krumm, S., Ziegler, M. & Buehner, M. (2008). Reasoning and working memory as predictors of school grades. *Learning and Individual Differences*, 18, 248-257.
- Kubinger, K. D. & Wurst, E. (2000). *Adaptives Intelligenz Diagnostikum 2 (AID 2)*. Göttingen: Beltz Test GmbH.
- Kun, A., Parsons, J. E. & Ruble, D. N. (1974). Development of integration processes using ability and effort information to predict outcomes. *Developmental Psychology*, 10, 721-732.
- 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(1), 148-161.
- Kyllonen, P. C. & Christal, R. E. (1990). Reasoning Ability Is (Little More Than) Working-Memory Capacity?! *Intelligence*, 14, 389-433.
- Laidra, K., Pullmann, H. & Allik, J. (2007). Personality and intelligence as predictors of academic achievement: A cross-sectional study from elementary to secondary school. *Personality and Individual Differences*, 42, 441-451.
- LaPointe, L. B. & Engle, R. W. (1990). Simple and Complex Word Spans as Measures of Working Memory Capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(6), 1118-1133.
- Lehmann, R. H., Peek, R. & Gänsfuß, R. (1997). *Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an Hamburger Schulen*. Hamburg: Behörde für Schule, Jugend und Berufsbildung, Amt für Schule.
- Lehto, J., Juujarvi, P., Kooistra, L. & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology*, 21, 59–80.

- Leventhal, T. & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin*, 126, 309-337.
- Levin, H. S., Fletcher, J. M., Kufera, J. A., Harward, H., Lilly, M. A., Mendelsohn, D. et al. (1996). Dimensions of cognition measured by the Tower of London and other cognitive tasks in head injured children and adolescents. *Developmental Neuropsychology*, 12, 17-34.
- Levpuscek, M. P & Zupancic, L. (2009). Math achievement in early adolescence: The role of parental involvement, teacher's behavior and student's motivational beliefs about math. *Journal of Early Adolescence*, 29, 541-570.
- Lienert, G. A. (1989). *Testaufbau und Testanalyse*. München: Psychologie Verlags Union.
- Lloyd, J. & Barenblatt, L. (1984). Intrinsic intellectuality: Its relations to social class, intelligence, and achievement. *Journal of Personality and Social Psychology*, 46, 646-654.
- Loehlin, J. C. (2000). Group differences in intelligence. In R. J. Sternberg (Hrsg.), *Handbook of intelligence* (S. 176- 193). Cambridge: Cambridge University Press.
- Logie, R. H. (1995). *Visuo-spatial working memory*. Hove, UK: Erlbaum.
- Logie, R. H. & Pearson, D. G. (1997). The inner eye and the inner scribe of visuo-spatial working memory: Evidence from developmental fractionation. *European Journal of Cognitive Psychology*, 9(3), 241-257.
- Lu, L., Weber, H. S., Spinath, F. M. & Shi, J. (2011). Predicting school achievement from cognitive and non-cognitive variables in a Chinese sample of elementary school children. *Intelligence*, 39, 130-140.
- Luo, D., Thompson, L. A. & Detterman, D. K. (2006). The criterion validity of tasks of basic cognitive processes. *Intelligence*, 34, 79-120.
- Lustig, C., May, C. & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General*, 130(2), 199-207.
- Magnusson, D. & Backtemann, G. (1978). Longitudinal stability of person characteristics: Intelligence and creativity. *Applied Psychological Measurement*, 2, 841-849.
- Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, 23, 129-149.
- Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology*, 79, 280-295.

- Marsh, H. W. (1990). Causal Ordering of Academic Self-Concept and Academic Achievement: A Multiwave, Longitudinal Panel Analysis. *Journal of Educational Psychology*, 82, 646-656.
- Marsh, H. W. & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. Hoyle (Hrsg.), *Structural equation modeling: Concepts, issues and applications* (S. 177–198). Thousand Oaks, CA: Sage.
- Marsh, H. W. & Hau, K. T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology*, 96, 56–67.
- Marsh, H. W., Byrne, B. M. & Yeung, S. A. (1999). Causal ordering of academic selfconcept and achievement: reanalysis of a pioneering study and revised recommendations. *Educational Psychologist*, 34, 155–167.
- Marsh, H. W., Kong, C.-K. & Hau, K.-T. (2000). Longitudinal Multilevel Models of the Big-Fish-Little-Pond Effect on Academic Self-Concept: Counterbalancing Contrast and Reflected-Glory Effects in Hong Kong Schools. *Journal of Personality and Social Psychology*, 78, 337-349.
- Marsh, H. W., Parker, J. W. & Smith, I. D. (1983). Preadolescent self-concept: Its relation to self-concept as inferred by teachers and to academic ability. *British Journal of Educational Psychology*, 53, 60–78.
- Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O. & Baumert, J. (2005). Academic Self-Concept, Interest, Grades, and Standardized Test Scores: Reciprocal Effects Models of Causal Ordering. *Child Development*, 76, 397-416.
- Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J. & Craven, R. G. (2008). The big-fish-little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychological Review*, 20(3), 319-350.
- McClelland, D. C., Atkinson, J. W., Clark, R. A & Lowell E. L. (1953). *The Achievement Motive*. Appleton- Century- Crofts.
- McGrew, K. S. (2005). The Cattell–Horn–Carroll theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Hrsg.), *Contemporary intellectual assessment: Theories, tests, and issues* (S. 136–181). New York: Guilford Press.
- McLean, J. F. & Hitch, J. (1999). Working memory impairments in children with specific arithmetical learning difficulties. *Journal of Experimental Child Psychology*, 74, 240–260.

- Meece, J. L., Blumenfeld, P. C. & Hoyle, R. H. (1988). Students' goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology*, 80, 514-523.
- Mehlhorn, G. & Mehlhorn, H.-G. (1981). *Intelligenz-Zur Erforschung und Entwicklung geistiger Fähigkeiten*. Berlin: Dt. Verlag der Wissenschaften.
- Miller, G. A., Galanter, E. & Pribram, K. H. (1960). *Plans and the structure of behavior*. New York: Holt, Rinehart and Winston.
- Miquelon, P. & Vallerand, R. J. (2008). Goal motives, well-being, and physical health: An integrative model. *Canadian Psychology*, 49, 241–249.
- Miyake, A. (2001). Individual differences in working memory: Introduction to the special section. *Journal of Experimental Psychology: General*, 130(2), 163-168.
- Miyake, A. & Shah, P. (1999). Toward Unified Theories of Working Memory. In A. Miyake & P. Shah (Hrsg.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (S. 442–481). Cambridge, UK: Cambridge University Press.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A. & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49–100.
- Moffitt, T. E., Caspi, A., Harkness, A. R. & Silva, P. A. (1993). The natural history of change in intellectual performance: Who changes? How much? Is it meaningful? *Journal of Child Psychology and Psychiatry*, 34, 455-506.
- Mortensen, E. L. & Kleven, M. (1993). A WAIS longitudinal study of cognitive development during the life span from ages 50 to 60. *Developmental Neuropsychology*, 9, 115-130.
- Moschner, B. & Dickhäuser, O. (2006). Selbstkonzept. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 83-94). Weinheim: Beltz Verlag.
- Müller-Benedict, V. (2007). Wodurch kann die soziale Ungleichheit des Schulerfolgs am stärksten verringert werden? *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 59, 615–639.
- Murphy, K. P. & Alexander, P. A. (2006). *Understanding How Students Learn. A Guide for Instructional Leaders*. Thousand Oaks: Corwin Press.
- Murray, H. A. (1938). *Explorations in Personality*. New York: Oxford University Press.
- Muthén, L. K. & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 4, 599–620.

- Neale, M. C. & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Dordrecht: Kluwer Academic Publications.
- Neale, M. C., Lubke, G., Aggen, S. H. & Dolan, C. V. (2005). Problems with using sum scores for estimating variance components: Contamination and measurement non-invariance. *Twin Research and Human Genetics*, 8, 553–568.
- Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J. et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51, 77-101.
- Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, 91(3), 328-346.
- Oberauer, K., Schulze, R., Wilhelm, O. & Süß, H. (2005). Working Memory and Intelligence – Their Correlation and Their Relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, 131(1), 61-65.
- Organisation for Economic Co-operation and Development (OECD). (Hrsg.). (2001). *Lernen für das Leben. Erste Ergebnisse der internationalen Schulleistungsstudie PISA 2000*. Paris: OECD.
- Oswald, H. & Krappmann, L. (2004). Soziale Ungleichheit in der Schulklasse und Schulerfolg. *Zeitschrift für Erziehungswissenschaften*, 7(4), 479-496.
- Passolunghi, M. C. & Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *Journal of Experimental Child Psychology*, 88, 348–367.
- Passolunghi, M. C., Mammarella, I. C. & Altoè, G. (2008). Cognitive abilities as precursors of the early acquisition of mathematical skills during first through second grades. *Developmental Neuropsychology*, 33(3), 329-250.
- Passolunghi, M. C., Vercelloni, B. & Schadee, H. (2007). The precursors of mathematics learning: Working memory, phonological ability and numerical competence. *Cognitive Development*, 22, 165–184.
- Paunonen, S. V. & Ashton, M. C. (2001). Big five predictors of academic achievement. *Journal of Research in Personality*, 35, 78–90.
- Pekrun, R. (2009). Emotions at school. In K. R. Wentzel & A. Wigfield (Hrsg.), *Handbook of motivation at school* (S. 575-604). New York: Routledge.
- Petrill, S. & Deater-Deckard, K. (2004). Task orientation, parental warmth and SES account for a significant proportion of the shared environmental variance in general cognitive ability in early childhood: evidence from a twin study. *Developmental Science*, 7(1), 25- 32.

- Petrill, S. A., Lipton, P. A., Hewitt, J. K., Plomin, R., Cherny, S. S., Corely, R., et al. (2004). Genetic and environmental contributions to general cognitive ability through the first 16 years of life. *Developmental Psychology*, 40, 805–812.
- Pickering, S. J., Gathercole, S. E. & Peaker, M. (1998). Verbal and visuo-spatial short-term memory in children: Evidence for common and distinct mechanisms. *Memory & Cognition*, 26, 1117–1130.
- Pickering, S. J., Gathercole, S. E., Hall, M. & Lloyd, S. (2001). Development of memory for pattern and path: Further evidence for the fractionation of visuo-spatial short-term memory. *Quarterly Journal of Experimental Psychology*, 54A, 397-420.
- Plomin, R., DeFries, J. C., McClearn, G. E. & McGuffin, P. (2008). *Behavioral Genetics* (Fifth Edition). New York: Worth Publishers.
- Postle, B. R., Ferrarelli, F., Hamidi, M., Feredoes, E., Massimini, M., Peterson, M., et al. (2006). Repetitive Transcranial Magnetic Stimulation dissociates working memory manipulation from retention functions in the prefrontal, but not Posterior Parietal, Cortex. *Journal of Cognitive Neuroscience*, 18, 1712–1722.
- Reinecke, J. (2005). *Strukturgleichungsmodelle in den Sozialwissenschaften*. München: Oldenbourg Verlag.
- Rheinberg, F. (2004a). *Motivation* (5. Auflage). Stuttgart: Kohlhammer.
- Rindermann, H. (2006). Was messen internationale Schulleistungsstudien? Schulleistungen, Schülerfähigkeiten, kognitive Fähigkeiten, Wissen oder allgemeine Intelligenz? *Psychologische Rundschau*, 57(2), 69-86.
- 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.
- Rohde, T. E. & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence*, 35(1), 83–92.
- Roth, P. L. & Clark, R. L. (1998). Meta-analyzing the relation between grades and salary. *Journal of Vocational Behavior*, 53, 386–400.
- Rowe, D. C., Vesterdal, W. J. & Rodgers, J. L. (1998). Herrnstein's syllogism: Genetic and shared environmental influences on IQ, education, and income. *Intelligence*, 26, 405-423.
- Rummer, R. & Fiebach, C. (2010). Prozessbasierte Ansätze in der aktuellen Arbeitsgedächtnisforschung. Editorial. *Psychologische Rundschau*, 61(1), 1-2.
- Ryan, R. M. & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development and well-being. *American Psychologist*, 55, 68-78.

- Saarländisches Ministerium für Bildung (2011). *Eckpunktepapier zur Gemeinschaftsschule*, 1-7. Abgerufen am 03.01.2012 unter <http://www.gemeinsam-bilden.de/mediathek>.
- Sauer, J. (2006). Prognose von Schulerfolg. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 83-94). Weinheim: Beltz Verlag.
- Schermelleh-Engel, K., Moosbrugger, H. & Müller, H (2003). Evaluating the Fit of Structural Equation Models: Test of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research Online*, 8, 23-74.
- Schick, A., Fehrenbach, C., Treutlein, A., Zöller, I., Roos, J. & Schöler, H. (2006). *EVES-Arbeitsberichte 6, Familiärer Hintergrund der Einschulungsjahrgänge 2001 und 2002 in Heidelberg. Sozioökonomischer Status, Bildungsnähe, Familienstruktur und außerschulische Förderung*. Heidelberg: Pädagogische Hochschule.
- Schicke, M. & Fagan, T. K. (1994). Contributions of self-concept and intelligence to the prediction of academic achievement among grade 4, 6, and 8 students. *Canadian Journal of School Psychology*, 10, 62–69.
- Schiefele, U. (1991). Interest, Learning and Motivation. *Educational Psychologist*, 26 (3&4), 299-323.
- Schiefele, U. (1999). Interest and learning from text. *Scientific Studies of Reading*, 3(3), 257-279.
- Schiefele, U. (2009). Motivation. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 151–177). Heidelberg: Springer.
- Schiefele, U. & Köller, O. (2006). Intrinsische und extrinsische Motivation. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (3., überarbeitete u. erw. Aufl., S. 303-310). Weinheim: Beltz.
- Schiefele, U. & Schreyer, I. (1994). Intrinsische Lernmotivation und Lernen. *Zeitschrift für Pädagogische Psychologie*, 8(1), 1-13.
- Schiefele, U. & Streblow, L. (2005). Intrinsische Motivation – Theorien und Befunde. In R. Vollmeyer & J. C. Brunstein (Hrsg.), *Motivationspsychologie und ihre Anwendung* (S. 39-58). Stuttgart: Kohlhammer.
- Schilling, S. R., Sparfeldt, J. R. & Rost, D. H. (2006). Facetten schulischen Selbstkonzepts – Welchen Unterschied macht das Geschlecht? *Zeitschrift für Pädagogische Psychologie*, 20, 9-18.
- Schlag, B. (2009). *Lern- und Leistungsmotivation*. Wiesbaden : VS Verl. für Sozialwiss.
- Schmidt-Atzert, L. & Deter, B. (1993). Intelligenz und Ausbildungserfolg: Eine Untersuchung zur prognostischen Validität des I-S-T-70. *Zeitschrift für Arbeits- und Organisationspsychologie*, 37, 52-63.

- Schmitt, N., Gooding, R. Z., Noe, R. A. & Kirsch, M. (1984). Metaanalyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*, 37, 407-422.
- Schneider, K. & Schmalt, H.-D. (2000). *Motivation* (3. Aufl.). Stuttgart: Kohlhammer.
- Schneider, W. (Hrsg.) (2008). *Entwicklung von der frühen Kindheit bis zum Erwachsenenalter. Befunde der Münchner Längsschnittstudie LOGIK*. Weinheim: Beltz.
- Schwartzman, A. E., Gold, D., Andres, D., Arbuckle, T. Y. & Chaikelson, J. (1987). Stability of intelligence: A 40-year follow-up. *Canadian Journal of Psychology*, 41, 244-256.
- Seginer, R. (2006). Parents' educational involvement: A developmental ecological perspective. *Parenting: Science and Practice*, 6, 1-48.
- Senn, T. E., Espy, K. A. & Kaufmann, P. M. (2004). Using path analysis to understand executive function organisation in preschool children. *Developmental Neuropsychology*, 26, 445-464.
- Shah, P. & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, 125(1), 4-27.
- Shah, P. & Miyake, A. (1999). Models of working memory: An introduction. In A. Miyake & P. Shah (Hrsg.), *Models of working memory: Mechanisms of active maintenance and executive control* (S. 1-27). New York: Cambridge University Press.
- Shavelson, R. J., Hubner, J. J. & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, 46, 407-441.
- Siegel, L. S. (1994). Working memory and reading: A life-span perspective. *International Journal of Behavioural Development*, 17, 109-124.
- Siegel, L. S. & Ryan, E. B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. *Child Development*, 60(4), 973-980.
- Skaalvik, E. M. & Skaalvik, S. (2004). Self-concept and self-efficacy: a test of the internal/external frame of reference model and predictions of subsequent motivation and achievement. *Psychological Reports*, 95, 1187-1202.
- Skaalvik, E. M., & Valas, H. (1999). Relations among achievement, self-concept, and motivation in mathematics and language arts: A longitudinal study. *Journal of Experimental Education*, 67, 135-149.
- Smith, E. E. & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology*, 33, 5-42.

- Snodgrass, J. G. & Vanderwart, M. (1980). A Standardized Set of 260 Pictures: Norms for Name Agreement, Image Agreement, Familiarity, and Visual Complexity. *Journal of Experimental Psychology*, 6(2), 174-215.
- Son, S.-H. & Morrison, F. J. (2010). The nature and impact of changes in home learning environment on development of language and academic skills in preschool children. *Developmental Psychology*, 46, 1103-1118.
- Spearman, C. (1904). 'General intelligence', objectively determined and measured. *American Journal of Psychology*, 15, 201-293.
- Spengler, M. (2009). *Differentielle Ätiologie der allgemeinen kognitiven Fähigkeit (g-Faktor). Vergleich einer deutschen und einer englischen Zwillingsstichprobe* [Differential Etiology of general cognitive ability (g factor): Comparison of a German and an English Twin Sample]. Dissertation at the Department of Psychology, Saarland University. URL: <http://scidok.sulb.uni-saarland.de/volltexte/2009/2631/>
- Spinath, B. (2005). Motivation als Kompetenz: Wie wird Motivation lehr- und lernbar? In R. Vollmeyer & J. C. Brunstein (Hrsg.), *Motivationspsychologie und ihre Anwendung* (S. 23-38). Stuttgart: Kohlhammer.
- Spinath, B. & Spinath, F. M. (2005). Longitudinal analysis of the link between learning motivation and competence beliefs among elementary school children. *Learning and Instruction*, 15, 87-102.
- Spinath, B. & Steinmayr, R. (2008). Longitudinal Analysis of Intrinsic Motivation and Competence Beliefs: Is There a Relation Over Time? *Child Development*, 79(5), 1555-1569.
- Spinath, F. M. & Wolf, H. (2006). CoSMoS and TwinPaW: Initial Report on two new German twin studies. *Twin Research and Human Genetics*, 9, 787-790.
- Spinath, B., Freudenthaler, H. & Neubauer, A. (2010). Predicting domain-specific school achievement in boys and girls by intelligence, personality and motivation. *Personality and Individual Differences*, 48, 481-486.
- Spinath, B., Spinath, F. M., Harlaar, N. & Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence*, 34, 363-374.
- Spinath, B., Stiensmeier-Pelster, J., Schöne, C. & Dickhäuser, O. (2002). *Die Skalen zur Erfassung von Lern- und Leistungsmotivation* (SELLMO) [Measurement scales for learning and performance motivation]. Göttingen: Hogrefe.
- St. Clair-Thompson, H. L. & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *The Quarterly Journal of Experimental Psychology*, 59(4), 745-759.

- Statistisches Bundesamt (Destatis) (Hrsg.). (2011). Datenreport 2011. Ein Sozialbericht für die Bundesrepublik Deutschland. Bonn: Bundeszentrale für politische Bildung. Abgerufen am 10.12.2011 unter <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Publikationen/Querschnittsveroeffentlichungen/Datenreport/Downloads/Datenreport2011Kap3,property=file.pdf>
- Stauffer, J., Ree, M. & Carreta, T. (1996). Cognitive-components tests are not much more than g: An extension of Kyllonen's analyses. *The Journal of General Psychology*, 123, 193–205.
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173–180.
- Steiger, J. H., Shapiro, A. & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika*, 50, 253–264.
- Steinkamp, M. W. & Maehr, M. L. (1983). Affect, ability, and science achievement: A quantitative synthesis of correlation research. *Review of Educational Research*, 53, 369–396.
- Steinmayr, R. & Spinath, B. (2007). Predicting school achievement from personality and motivation. *Zeitschrift für Pädagogische Psychologie*, 21, 207–216.
- Steinmayr, R. & Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning and Individual Differences*, 19, 80–90.
- Sternberg, R. J. (1997). The concept of intelligence and its role in lifelong learning and success. *American Psychologist*, 52, 1030–1037.
- Sternberg, R. J. & Detterman, D. K. (1986). *What is intelligence?* Norwood: Ablex Publishing Corporation.
- Stipek, D., Recchia, S. & Mc Clintic, S. (1992). Self-evaluation in young children. *Monographs of the society for research in child development*, 57(1).
- Süß, H.-M. (2001). Prädiktive Validität der Intelligenz im schulischen und außerschulischen Bereich. In E. Stern & J. Guthke (Hrsg.), *Perspektiven der Intelligenzforschung* (S. 109–135). Lengerich: Pabst.
- Süß, H., Oberauer, K., Wittmann, W. W., Wilhelm, O. & Schulze, R. (2002). Working-memory capacity explains reasoning ability – and a little bit more. *Intelligence*, 30, 261–288.
- Swanson, H. L. (1993). Working memory in learning disability subgroups. *Journal of Experimental Child Psychology*, 56, 87–114.
- Swanson, H. L. (1994). Short-term memory and working memory: Do both contribute to our understanding of academic achievement in children and adults with learning disabilities? *Journal of Learning Disabilities*, 27, 34–50.

- Swanson, H. L. (1996). Individual and age-related differences in children's working memory. *Memory & Cognition*, 24, 70–82.
- Swanson, H. L. (2008). Working memory and intelligence in children: What develops? *Journal of Educational Psychology*, 100(3), 581–602.
- Swanson H. L. & Berninger, V. (1995). The role of working memory in skilled and less skilled reader's comprehension. *Intelligence*, 21, 83–108.
- Swanson, L. & Kim, K. (2007). Working memory, short-term memory, and naming speed as predictors of children's mathematical performance. *Intelligence*, 35, 151–168.
- Tabachnick, B. G. & Fidell, L. S. (2001). *Using multivariate statistics*. Boston: Pearson A&B.
- Tabachnick, B. G. & Fidell, L. S. (2007). *Using Multivariate Statistics*. Boston: Pearson Education.
- Tent, L. (1998). Zensuren. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (S. 580-584). Weinheim: Beltz.
- Tent, L. (2006). Zensuren. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 873-880). Weinheim: Beltz.
- Thorndike, E. L. (1924). *Measurement of intelligence*. New York: Houghton Mifflin.
- Thurstone, L. L. (1938). *Primary and mental abilities*. Chicago: University of Chicago Press.
- Tillman, C. M., Nyberg, L. & Bohlin, G. (2008). Working memory components and intelligence in children. *Intelligence*, 36, 394-402.
- Towse, J. & Cowan, N. (2005). Working memory and its relevance for cognitive development. In W. Schneider, R. Schumann-Hengsteler & B. Sodian (Hrsg.), *Young children's cognitive development: Interrelationships among executive functioning, working memory, verbal ability, and theory of mind* (S. 9–37). Mahwah, NJ: Erlbaum.
- Towse, J. N., Hitch, G. J. & Hutton, U. (1998). A reevaluation of working memory capacity in children. *Journal of Memory and Language*, 39(2), 195-217.
- Turner, M. L. & Engle, R. W. (1989). Is Working Memory Capacity Task Dependent?. *Journal of Memory and Language*, 28, 127-154.
- Tzuriel, D. & Klein, P. S. (1983). Learning skills and types of temperaments as discrimination between intrinsically and extrinsically motivated children. *Psychological Reports*, 53, 59–69.

- Unsworth, N. (2010). On the division of working memory and long-term memory and their relation to intelligence: A latent variable approach. *Acta Psychologica*, 134, 16-28.
- Unsworth, N. & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. *Intelligence*, 33, 67-81.
- Unsworth, N. & Engle, R. W. (2007). The Nature of Individual Differences in Working Memory Capacity: Active Maintenance in Primary Memory and Controlled Search From Secondary Memory. *Psychological Review*, 114(1), 104-132.
- Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M. & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, 17, 635-654.
- Urdan, T. (2004). Predictors of academic self-handicapping and achievement: Examining achievement goals, classroom goal structures, and culture. *Journal of Educational Psychology*, 96, 251–264.
- Valentine, J. C., DuBois, D. L. & Cooper, H. (2004). The Relation between Self-Beliefs and Academic Achievement: A Meta-Analytic Review. *Educational Psychologist*, 39, 111-133.
- Vallar, G. & Papagno, C. (2002). Neuropsychological impairments of verbal short-term memory. In M. Kopelman, A. D. Baddeley & B. Wilson (Hrsg.), *Handbook of neuropsychology* (S. 249–270). Chichester, England: Wiley.
- van der Sluis, S., van der Leij, A. & de Jong, P. F. (2005). Working memory in Dutch children with reading- and arithmetic-related LD. *Journal of Learning Disabilities*, 38, 207–221.
- Vock, M. (2005). *Arbeitsgedächtniskapazität bei Kindern mit durchschnittlicher und hoher Intelligenz*. Dissertation. Universität Münster, Deutschland.
- Vock, M. & Holling, H. (2008). The measurement of visuo-spatial and verbal-numerical Working Memory: Development of IRT based scales. *Intelligence*, 36, 161-182.
- Vock, M., Preckel, F. & Holling, H. (2011). Mental Abilities and school achievement: A test of a mediation hypothesis. *Intelligence*, 39, 357-369.
- Vollmeyer, R. (2005). Einführung: Ein Ordnungsschema zur Integration verschiedener Motivationskomponenten. In R. Vollmeyer & J. C. Brunstein (Hrsg.), *Motivationspsychologie und ihre Anwendung* (S. 9-22). Stuttgart: Kohlhammer.

- Walker, C. O., Greene, B. A. & Mansell, R. A. (2006). Identification with academics, intrinsic/ extrinsic motivation, and self-efficacy as predictors of cognitive engagement. *Learning and Individual Differences*, 16, 1–12.
- Wang, M. C., Haertel, G. D. & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, 63, 249-294.
- Watermann, R. & Baumert, J. (2006): Entwicklung eines Strukturmodells zum Zusammenhang zwischen sozialer Herkunft und fachlichen und überfachlichen Kompetenzen: Befunde national und international vergleichender Analysen. In J. Baumert, P. Stanat & R. Watermann (Hrsg.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit. Vertiefende Analysen im Rahmen von PISA 2000* (S. 61-94). Wiesbaden.
- Weiber, R. & Mülhau, D. (2010). *Strukturgleichungsmodellierung. Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS*. Heidelberg: Springer.
- Weiner, B., (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, 92, 548–573.
- Weinert, F. E. & Helmke, A. (1995). Inter-classroom differences in instructional quality and interindividual differences in cognitive development. *Educational psychologist*, 30, 15-20.
- Weiß, H. R. (1998). *Grundintelligenztest Skala 2 mit Wortschatztest (WS) und Zahlenfolgetest (ZF) (CFT 20)*. Göttingen: Hogrefe.
- Weiß, R. H. (2006). CFT- 20- R. *Grundintelligenztest Skala 2*. Revision. Göttingen: Hogrefe.
- Welsh, M. C., Pennington, B. F. & Groisser, D. B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology*, 7, 131–149.
- West, S. G., Finch, J. F. & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In Hoyle, R. H. (Hrsg.), *Structural equation modeling: Concepts, issues, and applications* (S. 56-75). Thousand Oaks: Sage Publications.
- White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, 91, 461–481.
- White, R. H. (1959). Motivation reconsidered: the concept of competence. *Psychological Review*, 66, 297–333.
- Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review*, 6, 49-78.

- Wigfield, A. & Eccles, J. S. (1992). The Development of Achievement Task Values: A Theoretical Analysis. *Developmental Review*, 12, 265-310.
- Wigfield, A. & Eccles, J. S. (2000). Expectancy-Value Theory of Achievement Motivation. *Contemporary Educational Psychology*, 25, 68-81.
- Wigfield, A., Eccles, J. S. & Pintrich, P. R. (1996). Development between the ages of eleven and twenty-five. In D. C. Berliner & R. C. Calfee (Hrsg.), *The Handbook of Educational Psychology*. New York: Macmillan Publishing.
- Wigfield, A., Tonks, S. & Klauda, S. L. (2009). Expectancy-value theory. In K. R. Wentzel & A. Wigfield (Hrsg.), *Handbook of motivation in school* (S. 55–76). New York: Taylor Francis.
- Wild, K.-P. & Krapp, A. (1995). Elternhaus und intrinsische Lernmotivation. *Zeitschrift für Pädagogik*, 41(4), 579-595.
- Wilson, R. S. (1986). Continuity and change in cognitive ability profiles. *Behavior Genetics*, 16, 45-60.
- Yoon, C., Feinberg, F., Luo, T., Hedden, T., Gutches, A. H., Chen, H. M., Mikels, J. A., Jiao, S. & Park, D. C. (2004). A cross-culturally standardized set of pictures for younger and older adults: American and Chinese norms for name agreement, concept agreement, and familiarity. *Behavior Research Methods, Instruments, & Computers*, 36(4), 639-649.
- Zeidner, M. (1998). *Test anxiety: the state of the art*. New York: Plenum.
- Zusho, A., Pintrich, P. R. & Cortina, K. S. (2005). Motives, goals, and adaptive patterns of performance in Asian American and Anglo American students. *Learning and Individual Differences*, 15, 141–158.

Anhang A Tabellenanhang

Tabelle 22: Im Kinder- und Elternfragebogen erfasste Konstrukte mit jeweiliger Erhebungsperspektive

Konstrukt	Kinderfragebogen		Elternfragebogen	
	S	F	S	F
Motivation				
Erwartungs-Wert-Modell (Eccles)	+			+
Achievement Goals (Elliott)	+			+
Temperament				+
Involvement		+	+	
Erziehungsverhalten		+	+	
Normen und Werte			+	
SES			+	

Anmerkung: S: Selbstbericht; F: Fremdbbericht

Tabelle 23: Gängige "Daumenregeln" zur Bewertung der Modellgüte (in Anlehnung an Schermelleh-Engel et al., 2003)

Fit-Index	Guter Fit	Akzeptabler Fit
χ^2	$0 \leq \chi^2 \leq 2df$	$2df \leq \chi^2 \leq 3df$
p-Wert	$.05 < p \leq 1.0$	$.01 \leq p \leq .05$
χ^2/df	$0 \leq \chi^2/df \leq 2$	$2 < \chi^2/df \leq 3$
RMSEA	$0 \leq RMSEA \leq .05$	$.05 < RMSEA \leq .08$
Konfidenzintervall (CI)	Nahe RMSEA, linke Grenze des CI = 0	Nahe RMSEA
CFI	$.97 \leq CFI \leq 1.0$	$.95 \leq CFI \leq .97$
GFI	$.95 \leq CFI \leq 1.0$	$.90 \leq GFI \leq .95$
AIC	Kleiner als AIC des Vergleichsmodells	

Tabelle 24: Notation der Variablen im Pfadmodell der SEM (in Anlehnung an Weiber & Mülhhaus, 2010)

Notation	Bedeutung
η	Latente endogene Variable
ξ	Latente exogene Variable
y	Manifeste Indikatorvariable der endogenen Variable
x	Manifeste Indikatorvariable der exogenen Variable
ε	Fehlerterm der Messvariablen y
δ	Fehlerterm der Messvariable x
ζ	Fehlerterm der latenten endogenen Variable
β	Pfadgewicht zwischen zwei latenten endogenen Variablen
λ	Faktorladungen
γ	Pfadgewicht zwischen exogener und endogener Variable
$\longrightarrow$	Regressionsgewicht
$\longleftrightarrow$	Kovarianz bzw. Korrelation
	Latente Variable
	Manifeste Variable
	Fehlerterm

Anmerkung: vgl. Weiber & Mülhhaus (2010)

Tabelle 25: Vergleich der Faktorladungen und Korrelationen des Messmodells non-kognitiver Faktoren mit und ohne korrelierte Fehlerterme

	FSK1	FSK2	FSK3	FSK4	FSK5	FSK6	D	M
λ_{ohne}	.58	.65	.62	.62	.87	.76		
λ_{mit}	.59	.72	.54	.64	.84	.79		
	IM1	IM2	IM3	IM4	IM5	IM6		
λ_{ohne}	.59	.57	.61	.49	.87	.74		
λ_{mit}	.58	.61	.57	.50	.85	.76		
r_{ohne}							.83	.82
r_{mit}							.57	.74
r_{F1}							.42	.53
r_{F2}							.35	.50
r_{F3}							.51	.25

Anmerkungen: FSK1-6: FSK-Items, IM1-6: IM-Items; D: Deutsch; M: Mathematik; λ_{ohne} : Faktorladung im Messmodell ohne korrelierte Fehler; λ_{mit} : Faktorladung im Messmodell mit korrelierten Fehlern; r_{ohne} : Korrelation zwischen dem latenten FSK- und IM-Faktor ohne korrelierte Fehler; r_{mit} : Korrelation zwischen dem latenten FSK- und IM-Faktor mit korrelierten Fehlern; r_{F1} : Fehlerkorrelation 1; r_{F2} : Fehlerkorrelation 2; r_{F3} : Fehlerkorrelation 3.

Tabelle 26: Faktorladungen des jeweils parsimonischsten Modells der kognitiven Faktoren und SES

	Deutsch		Mathematik	
Modell	1a	3	2a	4a
DS	.63	.64	.64	.65
AS	.61	.59	.55	.57
MS	.42	.45	.45	.47
CFT1	.60	.62	.60	.61
CFT2	.45	.46	.46	.46
CFT3	.64	.64	.64	.64
CFT4	.51	.49	.52	.51
SES _M	.90	.87	.91	.90
SES _V	.70	.72	.69	.70
Note 1	.89	.89	.89	.90
Note 2	.96	.98	.87	.86

Anmerkungen: DS, AS, MS: Indikatoren des WM-Faktors; CFT1-4: Indikatoren des gf-Faktors; SES_M und SES_V: Indikatoren des SES-Faktors; Note 1 und 2: Indikatoren des Kriteriums. Modelle 1a/3: Modelle in Deutsch; 2a/4a: Modelle in Mathematik.

Tabelle 27: Faktorladungen des jeweils parsimonischsten Modells der kognitiven und nonkognitiven Faktoren, sowie SES

Modell	Deutsch		Mathematik	
	5f	7	6f	8a
DS	.64	.66	.66	.66
AS	.62	.59	.55	.58
MS	.43	.44	.47	.46
CFT1	.64	.65	.64	.65
CFT2	.42	.43	.42	.42
CFT3	.63	.63	.63	.63
CFT4	.49	.49	.46	.49
FSK1/4	.59	.59	.64	.65
FSK2/5	.69	.70	.80	.81
FSK3/6	.57	.57	.83	.82
IM1/4	.60	.57	.49	.50
IM2/5	.58	.59	.83	.84
IM3/6	.58	.58	.77	.76
SES _M	.81	.78	.89	.87
SES _V	.60	.63	.55	.56
Note 1	.91	.91	.88	.88
Note 2	.94	.94	.88	.88

Anmerkungen: DS, AS, MS: Indikatoren des WM-Faktors; CFT1-4: Indikatoren des gf-Faktors; FSK 1-6: Indikatoren des FSK-Faktors; IM1-6: Indikatoren des IM-Faktors; SES_M und SES_V: Indikatoren des SES-Faktors; Note 1 und 2: Indikatoren des Kriteriums. Modelle 5f/7: parsimonischste Modelle in Deutsch; 6f/8a: parsimonischste Modelle in Mathematik.

Tabelle 28: Korrelationen der untersuchten Konstrukte auf Item- bzw. Subskalenniveau

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	
1) CFT	.71	.64	.75	.68	.44	.37	.29	.29	.14	.04	.15	.12	.18	.22	.12	.11	-.08	-.12	.01	-.07	.11	.17	.08	.01	.26	.34	.45	
2) CFT1	-	.24	.40	.32	.36	.30	.23	.23	.18	.14	.14	.15	.15	.20	.08	.07	-.02	-.08	.02	-.01	.09	.15	.06	-.02	.16	.32	.36	
3) CFT2	-	-	.33	.29	.25	.20	.20	.15	.11	.07	.15	.03	.07	.07	.06	.02	-.06	-.05	.01	-.10	.00	.03	.02	-.06	.18	.20	.26	
4) CFT3	-	-	-	.30	.36	.28	.22	.27	.11	-.01	.09	.16	.14	.16	.09	.09	-.07	-.09	.01	-.07	.09	.14	.05	.04	.22	.27	.35	
5) CFT4	-	-	-	-	.25	.24	.16	.12	-.02	-.08	.04	-.02	.15	.17	.09	.11	-.08	-.12	-.05	-.02	.12	.14	.08	.06	.16	.15	.27	
6) WM	-	-	-	-	-	.78	.77	.61	.31	.20	.25	.25	.32	.26	.26	.29	.03	-.03	.07	.08	.16	.17	.13	.08	.15	.36	.45	
7) DS	-	-	-	-	-	-	.39	.24	.19	.13	.15	.16	.28	.23	.24	.24	.01	-.04	.06	.01	.13	.13	.14	.06	.12	.26	.37	
8) AS	-	-	-	-	-	-	-	.22	.33	.19	.30	.26	.19	.16	.16	.16	.09	.04	.07	.09	.06	.08	.05	.02	.11	.37	.33	
9) MS	-	-	-	-	-	-	-	-	.13	.11	.05	.12	.23	.19	.17	.24	-.04	-.08	.02	-.05	.17	.17	.12	.11	.09	.14	.26	
10) FSKg	-	-	-	-	-	-	-	-	.74	.79	.77	.77	.29	.29	.22	.26	.50	.39	.35	.42	.07	.03	.09	.06	.18	.63	.39	
11) FSK1	-	-	-	-	-	-	-	-	-	.45	.33	.10	.12	.10	.10	.09	.33	.45	.16	.18	-.06	-.08	-.01	-.05	.18	.45	.24	
12) FSK2	-	-	-	-	-	-	-	-	-	-	.38	.29	.31	.20	.23	.23	.39	.23	.42	.24	.07	.06	.07	.04	.17	.53	.35	
13) FSK3	-	-	-	-	-	-	-	-	-	-	-	.27	.23	.19	.26	.43	.25	.22	.22	.52	.13	.06	.12	.12	.08	.46	.31	
14) FSKm	-	-	-	-	-	-	-	-	-	-	-	-	-	.82	.87	.85	.10	-.02	.14	.11	.66	.45	.61	.53	.18	.37	.56	
15) FSK4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.53	.52	.12	.12	.01	.14	.10	.53	.56	.39	.32	.15	.29	.45
16) FSK5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.69	.06	-.04	.12	.05	.60	.28	.67	.50	.17	.28	.46	
17) FSK6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.09	-.01	.09	.11	.55	.29	.48	.55	.15	.38	.52	
18) IMg	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.73	.78	.80	.22	.14	.16	.23	-.04	.27	.04	
19) IM1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.36	.38	.02	-.02	-.01	.08	.03	.23	.01	
20) IM2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.42	.22	.17	.19	.17	-.03	.20	.05	
21) IM3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.24	.15	.16	.26	-.09	.20	.04	
22) IMm	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.74	.87	.81	.01	.13	.31	
23) IM4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.43	.34	.00	.07	.21	
24) IM5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.65	.03	.14	.30	
25) IM6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.00	.12	.24	
26) SES	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.32	.32	
27) Deutsch	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.73	
28) Mathe	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	

Anmerkung: CFT: CFT-Faktor; CFT1-4: CFT Subtests; WM: WM-Faktor; DS-MS: WM-Tasks; FSKg: Skalenwert FSK in Deutsch; FSK1-3: FSK-Items Deutsch; FSKm: Skalenwert FSK in Mathematik; FSK4-6: FSK-Items Mathematik; IMg: Skalenwert IM in Deutsch; IM1-3: IM-Items Deutsch; IMm: Skalenwert in Mathematik; IM4-6: IM-Items in Mathematik; SES: Höchster Bildungsabschluss der Eltern; Deutsch: Notendurchschnitt Deutsch; Mathe: Notendurchschnitt Mathe. Alle r signifikant mit $p < .05$, außer *kursiv* = nicht signifikant.

Anhang B Materialanhang

Darbietungsliste Digit Sorting

1. 1 - 5 - 3
2. 11 - 8
3. 4 - 1 - 9 - 3 - 8
4. 6 - 11 - 10 - 5 - 4 - 2
5. 9 - 12 - 6
6. 3 - 2 - 5 - 4
7. 9 - 10 - 2 - 11 - 5
8. 6 - 4 - 9 - 11 - 10 - 12
9. 11 - 10
10. 9 - 6 - 8 - 4

Darbietungsliste Animal Sorting

1. Schwein, Pferd
2. Spinne, Fisch, Schwein, Pferd, Giraffe
3. Schnecke, Fisch, Katze, Schwein, Pferd, Giraffe
4. Fisch, Schwein, Kamel
5. Fisch, Katze, Schwein, Giraffe
6. Schnecke, Fisch, Ente, Katze, Kamel
7. Maus, Ente, Hund, Pferd
8. Maus, Ente
9. Spinne, Hund, Pferd
10. Spinne, Maus, Ente, Katze, Hund, Kamel

Darbietungsliste Backward Matrix Span

1

1		2

2

		2
1		

3

	1	
		3
2		

4

	2	
1		3
	4	

5

1	3	
	5	
2		4

6

3		1
	2	

7

1		2
	4	
		3

8

1		2
3	4	
5		6

9

	1	
3		4
	5	2

10

4		6
1	2	3
	5	

Lehrerformular zur Erfassung der Schulnote mit schematischer Darstellung des Notenformulars

Liebe Lehrerin, lieber Lehrer,

im Rahmen unserer Studie bitten wir Sie, die Notentabelle für die getesteten Schüler folgendermaßen auszufüllen:

Während der Testung

- Bitte tragen Sie die Vor- und Nachnamen der Schüler zu den anonymen Zahlencodes ein, damit nach der Testung eine sichere und fehlerfreie Zuordnung der Noten zu den einzelnen Schülern möglich ist.

Nach der Testung

- Bitte tragen Sie für jeden Schüler die Jahres- und Halbjahreszeugnisnoten der 3. bzw. 4. Klasse in den Fächern Deutsch (D) und Mathematik (M) ein.
- Bitte kennzeichnen Sie den Rücklauf der Fragebögen in der vorgesehenen Spalte.
- Bitte trennen Sie bei Rückgabe des Notenformulars an die Testleitung die Namensspalte ab, um die Anonymisierung der Daten zu gewährleisten!

Vielen Dank für Ihre Unterstützung!

Vor- und Nachname	Zahlencode	Rückgabe	Jahreszeugnis 3. Klasse		Halbjahreszeugnis 4. Klasse	
		Fragebogen [ja/nein]	D	M	D	M
Max Mustermann	11111					
....	22222					
.....	33333					

Zunächst würden wir gerne wissen, wie gerne du gewisse Dinge hast. Versuche dazu ganz allgemein zu überlegen, wie sehr du die folgenden Dinge **magst**.

mag es überhaupt
nicht
mag es nicht
weder noch
mag es
mag es
sehr

1. ...Lesen?	☐	☐	☐	☐	☐
2. ...Schreiben (z.B. Aufsätze)?	☐	☐	☐	☐	☐
3. ...Rechtschreibung?	☐	☐	☐	☐	☐
4. ...Zahlenrätsel oder Textaufgaben lösen?	☐	☐	☐	☐	☐
5. ...Kopfrechnen?	☐	☐	☐	☐	☐
6. ...Multiplizieren und Dividieren?	☐	☐	☐	☐	☐

Jetzt sollst du bitte einschätzen, wie gut du Dinge deiner Meinung nach kannst. Es geht zwar wieder um dieselben Dinge wie eben, aber bitte sag uns dieses Mal was du glaubst, wie **gut du** in diesen Dingen bist!

	<i>gar nicht gut</i>	<i>nicht so gut</i>	<i>weder noch</i>	<i>ziemlich gut</i>	<i>sehr gut</i>
Wie gut bist du...					
1. ... im Lesen?	☐	☐	☐	☐	☐
2. ... im Schreiben (z.B. Aufsätze)?	☐	☐	☐	☐	☐
3. ... in Rechtschreibung?	☐	☐	☐	☐	☐
4. ... im Lösen von Zahlenrätseln oder Textaufgaben?	☐	☐	☐	☐	☐
5. ... im Kopfrechnen?	☐	☐	☐	☐	☐
6. ... im Multiplizieren und Dividieren?	☐	☐	☐	☐	☐
7. ... in der Schule?	☐	☐	☐	☐	☐

Item zur Erfassung des höchsten Bildungsabschlusses der Eltern
(Elternfragebogen)

1. Höchster Schul-/ Studienabschluss der Eltern:

	Mutter	Vater
unbekannt	0	0
ohne Schulabschluss	0	0
mit der letzten Hauptschulklasse abgeschlossen	0	0
Real-(Mittel-) oder Handelsschule ohne Abschlussprüfung	0	0
Real-(Mittel-) oder Handelsschule mit Abschlussprüfung	0	0
Gymnasium (höhere Schule) ohne Abitur	0	0
Abitur ohne anschließendes Studium	0	0
Abitur mit (noch) nicht abgeschlossenem Studium	0	0
Abitur mit abgeschlossenem Studium	0	0

Kategorienbildung:

1. ohne Angabe: unbekannt
2. ohne Abschluss: ohne Abschluss
3. Hauptschulabschluss: mit der letzten Hauptschulklasse abgeschlossen/ Real-(Mittel-) oder Handelsschule ohne Abschlussprüfung
4. Realschulabschluss: Real-(Mittel-) oder Handelsschule mit Abschlussprüfung/ Gymnasium (höhere Schule) ohne Abitur
5. Fachabitur/Abitur: Abitur ohne anschließendes Studium/ Abitur mit (noch) nicht abgeschlossenem Studium
6. Studium (in SEM-Analysen): Abitur mit abgeschlossenem Studium

Anhang C

Lebenslauf

Heike Sabine Weber

Persönliche Informationen

Staatsangehörigkeit:	Deutsch
Geburtsdatum/-ort:	25.08.1982/ Neunkirchen
Telefon:	0681 302 3038
E-Mail:	h.weber@mx.uni-saarland.de

Berufliche Tätigkeiten

Seit 02/2005:	INSTITUT FÜR BERUF UND BILDUNG (IBB), Saarbrücken/Neunkirchen Freie Mitarbeiterin
06/2008-05/2011:	UNIVERSITÄT DES SAARLANDES, LEHRSTUHL FÜR DIFFERENTIELLE PSYCHOLOGIE UND PSYCHOLOGISCHE DIAGNOSTIK, Saarbrücken Promotionsstipendium im internationalen Graduiertenkolleg IRTG Adaptive Minds

06/2011-05/2012: UNIVERSITÄT DES SAARLANDES, LEHRSTUHL
FÜR DIFFERENTIELLE PSYCHOLOGIE UND
PSYCHOLOGISCHE DIAGNOSTIK,
**Stipendium der Landesgraduiertenförderung der
Universität des Saarlandes**

Studium

10/2001 – 07/2002: TECHNISCHE UNIVERSITÄT KAISERSLAUTERN,
Kaiserslautern
Diplomstudiengang: **Lebensmittelchemie**

10/2002 – 04/2008: UNIVERSITÄT DES SAARLANDES, Saarbrücken
Diplomstudiengang: **Psychologie**
- 04/2008: Diplom (Abschlussnote: 1,1)

Sonstige Praxiserfahrung

09/2005 - 10/2005: SHG KLINIKEN, ARBEITS-THERAPIE-ZENTRUM
(ATZ), Saarbrücken
Praktikantin

04/2005 - 04/2008: UNIVERSITÄT DES SAARLANDES, LEHRSTUHL FÜR
KLINISCHE PSYCHOLOGIE UND PSYCHOTHERAPIE,
Saarbrücken
Studentische Hilfskraft

2006: GNOSIS FACIALIS, Institut für Verhaltens- und
Mimikforschung, Saarbrücken
Forschungspraktikantin

04/2008 - 06/2008: GERMAN SWISS INTERNATIONAL SCHOOL,
Hong Kong, China
Praktikantin

Publikationen

Hochschulschriften:

Einzelfallanalyse des Flirtverhaltens einer Patientin mit Posttraumatischer Belastungsstörung in der Therapie (April, 2008), Diplomarbeit.

Internationale Publikationen (peer-reviewed):

Lu, L., **Weber, H.S.**, Spinath, F.M. & Shi, J. (2011). Predicting school achievement from cognitive and non-cognitive variables in a Chinese sample of elementary school children. *Intelligence*, 39, 130-140.

Konferenzbeiträge und Präsentationen:

Weber, H.S. & Spinath, F.M. (2009). Predicting school success – the influence of intelligence, working memory and motivation. Presentation at the 14th Biennial Meeting of the International Society for the Study of Individual Differences, Evanston, Chicago, USA. 18. – 22.07.2009

Weber, H.S. & Spinath, F.M. (2009). Welche Rolle spielen Intelligenz, Arbeitsgedächtniskapazität und Motivation bei der Vorhersage von Schulerfolg? Vortrag bei der 10. Fachgruppentagung Differentielle Psychologie, Persönlichkeitspsychologie und psychologische Diagnostik, Landau, Deutschland. 28.-30.09.2009

Weber, H.S. & Spinath, F.M. (2010). Predicting school success: the role of working memory, intelligence and motivation. Gastvortrag am Institute of Psychology der Chinese Academy of Sciences, Beijing.

Weber, H.S., Lu, L., Spinath, F.M., & Shi, J. (2011). The role of working memory capacity, intelligence, and motivation in the prediction of school success: a cross-

cultural comparison. Presentation at the 15th Biennial Meeting of the International Society for the Study of Individual Differences (ISSID), London, July 25-29, UK.

Weber, H.S., Lu, L., Spinath, F.M., & Shi, J. (2011). Welche Rolle spielen Arbeitsgedächtniskapazität, Intelligenz und Motivation bei der Vorhersage von Schulerfolg? Ein deutsch-chinesischer Vergleich. Vortrag bei der 11. Fachgruppentagung Differentielle Psychologie, Persönlichkeitspsychologie und psychologische Diagnostik, Saarbrücken, Deutschland. 26.- 28.09.2011

Fremdsprachenkenntnisse

Englisch: fließend in Wort und Schrift

Französisch: Grundkenntnisse in Wort und Schrift

Chinesisch: Grundkenntnisse in Wort und Schrift

EDV-Kenntnisse

Anwendungen: Fortgeschrittene Kenntnisse in MS Office (Word, Powerpoint, Excel)

Statistiksoftware: Fortgeschrittene Kenntnisse in SPSS und AMOS

Besondere Kenntnisse

Zertifizierte FACS-Userin (Facial Action Coding System; Ekman & Friesen (1986))