
**Eine Multi-Kohorten-Sequenz-Analyse des
Zusammenhangs von Motivation und
Schulerfolg unter Berücksichtigung
allgemeiner kognitiver Fähigkeiten**

Ergebnisse aus der Zwillingsstudie KoSMoS

Dissertation
zur Erlangung des akademischen Grades eines

Doktors der Philosophie

der Philosophischen Fakultät III
der Universität des Saarlandes

vorgelegt von

Juliana Gottschling

aus Ottweiler

Saarbrücken, 2013

Dekan: Prof. Dr. Roland Brünken

Berichtersteller/innen:

Prof. Dr. Frank M. Spinath

Prof. Dr. Birgit Spinath

Prof. Dr. Gisa Aschersleben

Tag der Disputation: 14. Februar 2013

Danksagung

Mein Dank an dieser Stelle gilt all denen, die durch ihre vielfältige Unterstützung wesentlich zum Gelingen dieser Arbeit beigetragen haben.

Mein erster Dank gilt Prof. Dr. Frank M. Spinath für seine Unterstützung bei meiner Arbeit und die Freiheit, dabei auch eigene Wege gehen zu können. Meinen Kollegen möchte ich für den regen fachlichen Austausch, aber auch die Ablenkung und freundschaftliche Unterstützung in den letzten Jahren danken: Ralph, danke für das richtige Lied zur richtigen Zeit und die Erweiterung der Büroskala auf 15 und danke Gundula für die „Montagssitzung“ und so vieles mehr!

Weiterhin möchte ich den vielen Hilfskräften für ihre tatkräftige Unterstützung bei der Datenerhebung und -eingabe sowie Olga Dragunowa für ihre wertvolle und stets zuverlässige Arbeit danken. Natürlich wäre diese Arbeit ohne das Engagement der vielen Zwillingfamilien über einen Zeitraum von sechs Jahren hinweg nicht möglich gewesen, ihnen gilt daher ein besonderer Dank.

Heike Weber und Heike Maas möchte ich für das intensive Korrekturlesen der Arbeit und die vielen hilfreichen Anmerkungen herzlich danken sowie Elisabeth Hahn, Corinna Reichl und Gundula Stoll für ihre Flexibilität und spontane Lesebereitschaft. Für die vielen konstruktiven Diskussionen der verhaltensgenetischen Daten gilt mein Dank insbesondere Elisabeth Hahn sowie Marion Spengler und Heike Maas.

Ein großer Dank geht auch an meine Familie, die mir durch ihr Verständnis und ihre Zuversicht immer eine große Unterstützung sind. Eva Hoffmann und Sandra Strohm möchte ich für eine Freundschaft danken, die alles übersteht und keiner großen Worte bedarf.

Meinem Mann, Jürgen Rachor, danke ich von ganzem Herzen für seinen grenzenlosen Optimismus, seine Nervenstärke und die Selbstverständlichkeit von Dingen, die nicht selbstverständlich sind!

Inhaltsverzeichnis

Abbildungsverzeichnis	VIII
Tabellenverzeichnis	X
Abkürzungsverzeichnis	XI
I Einleitung	1
II Theoretische Grundlagen und empirische Befunde	7
1 Das Kriterium Schulerfolg	9
1.1 Operationalisierung	9
1.2 Entwicklung schulischen Erfolgs	10
2 Bedingungsfaktoren schulischen Erfolgs: Allgemeine kognitive Fähigkeit	15
2.1 Begriffsbestimmung	16
2.2 Strukturmodelle der allgemeinen kognitiven Fähigkeit	17
2.2.1 Die Generalfaktorentheorie von Charles Spearman	17
2.2.2 Das Three-Stratum-Modell von John B. Carroll	18
2.2.3 Abschließende Bewertung	19
2.3 Vorhersage von Schulerfolg durch allgemeine kognitive Fähigkeiten	20
2.4 Individuelle Entwicklungsstabilität allgemeiner kognitiver Fähigkeiten	22
2.5 Abschließende Bewertung	24
3 Bedingungsfaktoren schulischen Erfolgs: Motivation	27
3.1 Begriffsbestimmung	28
3.2 Das Konstrukt Leistungsmotivation	30
3.3 Das Erwartungs-Wert-Modell der Leistungsmotivation sensu Eccles	32
3.3.1 Erwartungskomponente: Fähigkeitsselbstkonzept	34
3.3.2 Wertkomponente: Intrinsische Motivation	36
3.3.3 Abschließende Bewertung	36
3.4 Vorhersage von Schulerfolg durch motivationale Variablen	37
3.4.1 Prognostische Relevanz des Fähigkeitsselbstkonzeptes	38
3.4.2 Prognostische Relevanz intrinsischer Motivation	43
3.4.3 Relative Bedeutung motivationaler Faktoren unter Berücksichtigung allgemeiner kognitiver Fähigkeiten	46
3.5 Individuelle Entwicklungsstabilität motivationaler Variablen	49

3.5.1	Individuelle Entwicklungsstabilität des Fähigkeitsselbstkonzeptes	50
3.5.2	Individuelle Entwicklungsstabilität intrinsischer Motivation	53
3.6	Abschließende Bewertung	55
4	Der Ansatz der Verhaltensgenetik zur Erklärung der Ursachen interindividueller Unterschiede	59
4.1	Grundlagen der Verhaltensgenetik	59
4.2	Abgrenzung von Anlage und Umwelt – Der Begriff der Erbllichkeit	62
4.3	Varianzzerlegung	63
4.3.1	Genetische Varianz	64
4.3.2	Umweltvarianz	65
4.4	Das klassische Zwillingsdesign als Methode der quantitativen Verhaltensgenetik	68
4.5	Schätzung von genetischer Varianz und Umweltvarianz	72
4.6	Befunde verhaltensgenetischer Studien	74
4.6.1	Verhaltensgenetische Studien zu Schulerfolg	74
4.6.2	Verhaltensgenetische Studien zu allgemeinen kognitiven Fähigkeiten	79
4.6.3	Verhaltensgenetische Studien zur Motivation	85
4.6.4	Verhaltensgenetische Studien zur gemeinsamen Betrachtung schulischen Erfolgs und seiner Prädiktoren	88
4.7	Abschließende Bewertung	89
III	Fragestellungen und Hypothesen	91
5	Fragestellung und Hypothesen der phänotypischen Betrachtung	93
5.1	Intra- und Interindividuelle Stabilität der Variablen	93
5.2	Längsschnittliche Analyse der Vorhersage von Schulerfolg	95
6	Fragestellung und Hypothesen der ätiologischen Betrachtung	99
6.1	Univariate Betrachtung	99
6.2	Multivariate Betrachtung	100
6.2.1	Ursachen interindividueller Stabilität	100
6.2.2	Verhaltensgenetische cross-lagged Betrachtung der Vorhersage von SE	101
IV	Datengrundlage und methodisches Vorgehen	103
7	Stichprobe	105
7.1	Die Zwillingsstudie zu Einflüssen von kognitiven Fähigkeiten und selbst eingeschätzter Motivation auf Schulerfolg (KoSMoS)	105
7.2	Auswahl der Untersuchungsstichprobe	107
7.2.1	Drop-out-Analyse	109
7.2.2	Stichproben der phänotypischen und ätiologischen Analysen	113
8	Erhebungsinstrumente	117
8.1	Zygotie	117

8.2	Erfassung des Kriteriums Schulnoten	118
8.3	Kognitive Fähigkeiten	119
8.3.1	Vergleich von Heim- und Telefon- testung kognitiver Fähigkeiten	119
8.3.2	Die kognitive Testbatterie	121
8.4	Motivation	124
8.4.1	Fähigkeitsselbstkonzept	124
8.4.2	Intrinsische Motivation	124
9	Analysen der phänotypischen Betrachtung	125
9.1	Voraussetzungen der Anwendbarkeit der SEM	127
9.2	Modellschätzung und Güteprüfung	131
9.3	SEM im Rahmen längsschnittlicher Analysen	136
10	Analysen der ätiologischen Betrachtung	143
10.1	Univariate genetische Analysen	143
10.2	Multivariate genetische Analysen	146
V	Ergebnisse	153
11	Deskriptive Statistik und manifeste Skaleninterkorrelationen	155
11.1	Schulerfolg: Kennwerte und deskriptive Statistik	155
11.2	Allgemeine kognitive Fähigkeiten: Kennwerte und deskriptive Statistik	158
11.3	Motivation: Kennwerte und deskriptive Statistik	165
12	Güteprüfung der latenten Messmodelle und latente Zusammenhänge	171
12.1	Voraussetzungsprüfung	171
12.2	Modell-Fit der Messmodelle und Invarianzprüfung	174
12.2.1	Messmodell des schulischen Erfolgs	174
12.2.2	Messmodell der kognitiven Fähigkeiten	175
12.2.3	Messmodell von FSK und IM	177
12.2.4	Überprüfung der Messinvarianz über die Zeit	179
13	Ergebnisse zur Beantwortung der phänotypischen Fragestellungen	183
13.1	Intra- und Interindividuelle Stabilität	183
13.1.1	Intraindividuelle Stabilität	183
13.1.2	Interindividuelle Stabilität	188
13.2	Längsschnittlicher Zusammenhang zwischen den Variablen	190
13.2.1	Motivation und Schulerfolg	190
13.2.2	Allgemeine kognitive Fähigkeiten, Motivation und Schulerfolg	194
13.3	Vorhersage des Schulerfolgs im MKSD	195
14	Ergebnisse zur Beantwortung der ätiologischen Fragestellungen	213
14.1	Univariate verhaltensgenetische Analysen	213
14.2	Multivariate verhaltensgenetische Analysen	221

14.2.1 Ursachen interindividueller Stabilität	221
14.2.2 Verhaltensgenetische Cross-Lagged Analyse	230
VI Diskussion	233
15 Diskussion der Ergebnisse aus der phänotypischen Betrachtung	235
15.1 Intra- und interindividuelle Stabilität	235
15.1.1 Schulerfolg in Mathematik und Deutsch	235
15.1.2 Allgemeine kognitive Fähigkeiten	240
15.1.3 Fähigkeitsselbstkonzept und Intrinsische Motivation	244
15.2 Vorhersage schulischen Erfolgs	249
15.2.1 Latente Interkorrelation der Konstrukte	250
15.2.2 Vorhersage schulischen Erfolgs innerhalb des MKSD	255
15.2.3 Inkrementelle Validität des FSKs Mathematik	265
16 Ergebnisse der ätiologischen Betrachtung	267
16.1 Univariate und multivariate genetische Analyse schulischen Erfolgs	268
16.2 Univariate genetische Analyse allgemeiner kognitiver Fähigkeiten	273
16.3 Univariate und multivariate genetische Analyse von FSK und IM	274
16.4 Verhaltensgenetische Cross-Lagged Analyse	278
17 Fazit	281
17.1 Abschließende Zusammenfassung und Implikationen der Studie	281
17.2 Limitationen	285
17.3 Ausblick	290
A Abbildungsanhang	293
B Tabellenanhang	297
C Materialanhang	307
Literaturverzeichnis	311

Abbildungsverzeichnis

Abb. 2.1	Das Three-Stratum-Modell von Carroll	19
Abb. 3.1	Erwartungs-Wert-Modell der Leistungsmotivation nach Eccles . .	32
Abb. 3.1	Prototypisches Modell zur Überprüfung der kausalen Wirkrichtung zwischen Fähigkeitsselbstkonzept und schulischem Erfolg	41
Abb. 4.1	Erblichkeit schulischen Erfolgs	77
Abb. 4.2	Erblichkeit der Intelligenz	81
Abb. 7.1	Beschreibung der Gesamtstichprobe	108
Abb. 7.2	Verbleib der Stichprobe über den Untersuchungsverlauf hinweg . .	110
Abb. 7.3	Höchster Bildungsabschluss im Vergleich mit einer repräsentativen deutschen Vergleichsstichprobe	111
Abb. 7.4	Besuchte Schulform der Kinder im Vergleich mit einer deutschen Vergleichsstichprobe	112
Abb. 7.5	Klassenstufe der Kinder über den Untersuchungsverlauf	114
Abb. 9.1	Schematische Darstellung eines cross-lagged panel Designs	137
Abb. 9.2	Cross-lagged panel Design mit Kontrolle der Autokorrelation . . .	138
Abb. 9.3	Prototypischen Multi-Kohorten-Sequenz-Design	140
Abb. 10.1	Univariates latentes Pfaddiagramm	144
Abb. 10.1	Bivariate latente Cholesky Zerlegung	148
Abb. 10.2	Genetisches Cross-Lagged Modell	151
Abb. 13.1	Full-Forward Modell Fähigkeitsselbstkonzept Mathematik	197
Abb. 13.2	Multi-Kohorten Modell Fähigkeitsselbstkonzept Mathematik	198
Abb. 13.3	Full-Forward Modell Fähigkeitsselbstkonzept Deutsch	202
Abb. 13.4	Multi-Kohorten Modell Fähigkeitsselbstkonzept Deutsch	203
Abb. 13.5	Full-Forward Modell Intrinsische Motivation Mathematik	206
Abb. 13.6	Multi-Kohorten Modell Intrinsische Motivation Mathematik	207
Abb. 13.7	Full-Forward Modell Intrinsische Motivation Deutsch	209
Abb. 13.8	Multi-Kohorten Modell Intrinsische Motivation Deutsch	210
Abb. 13.9	Multi-Kohorten Modell Fähigkeitsselbstkonzept Mathematik mit Kontrolle von g	211
Abb. 14.1	Parameterschätzung der latenten univariaten vollständigen Modelle	215
Abb. 14.1	Cholesky Zerlegung des schulischen Erfolgs in Mathematik	223
Abb. 14.2	Cholesky Zerlegung des schulischen Erfolgs in Deutsch	224
Abb. 14.3	Cholesky Zerlegung des Fähigkeitsselbstkonzeptes Mathematik . .	227
Abb. 14.4	Cholesky Zerlegung der intrinsischen Motivation Mathematik . . .	228

Abb. 14.5	Cholesky Zerlegung des Fähigkeitsselbstkonzeptes Deutsch	229
Abb. 14.6	Cholesky Zerlegung der intrinsischen Motivation Deutsch	230
Abb. 1.1	Darstellung von Mess-und Strukturmodell der Strukturgleichungs- analyse	293
Abb. 1.2	Messmodell Fähigkeitsselbstkonzept und intrinsische Motivation . .	293
Abb. 1.3	Messmodell der allgemeinen kognitiven Fähigkeiten	294
Abb. 1.4	Messmodell Schulerfolg	294
Abb. 1.5	Multivariates verhaltensgenetisches Modell	295

Tabellenverzeichnis

4.1	Genetische Korrelationen und Umweltkorrelationen in verschiedenen Verwandtschaftsbeziehungen	65
7.1	Beschreibung der Untersuchungsstichprobe K_AS zu Messzeitpunkt 1	109
7.2	Beschreibung der Analysestichprobe der phänotypischen Analysen zu Messzeitpunkt 1	113
7.3	Zygotiegruppen der Analysestichproben der ätiologischen Analysen	115
11.1	Deskriptive Kennwerte und Verteilungsmaße der Schulnoten	157
11.2	Interkorrelation und Reliabilität der Schulnoten (Mathematik und Deutsch) innerhalb und zwischen den Schuljahren	158
11.3	Deskriptive Kennwerte und Verteilungsmaße der kognitiven Skalen	160
11.4	Reliabilitäten, Interkorrelationen und Faktorladungen der kognitiven Messung	162
11.5	Manifeste Interkorrelation zwischen dem g -Faktor und Schulnoten	164
11.6	Deskriptive Kennwerte und Verteilungsmaße der Motivationsskalen	166
11.7	Interkorrelation und Reliabilität der Motivationsskalen	168
11.8	Manifeste Interkorrelation zwischen g -Faktor und Motivation	169
12.1	Modell-Fit des Messmodells – Schulerfolg Mathematik und Deutsch	176
12.2	Modell-Fit des Messmodells – Kognitive Fähigkeiten	178
12.3	Modell-Fit des Messmodells – FSK Mathematik und Deutsch	180
12.4	Gütemodelle – Motivation – IM Mathematik und Deutsch	181
13.1	Modellprüfung der Veränderung der latenten Mittelwerte und Varianzen – Schulerfolg Mathematik und Deutsch	185
13.2	Modellprüfung der Veränderung der latenten Mittelwerte und Varianzen – FSK Mathematik und Deutsch	186
13.3	Modellprüfung der Veränderung der latenten Mittelwerte und Varianzen – IM Mathematik und Deutsch	187
13.4	Latente Mittelwerte und Varianzen von Schulerfolg, Fähigkeitsselbstkonzept und intrinsischer Motivation	188
13.5	Manifeste und latente Stabilität von Schulerfolg, Fähigkeitsselbstkonzept und intrinsischer Motivation	189
13.6	Latente Skaleninterkorrelation zwischen Schulerfolg und Motivation	191
13.7	Latente Skaleninterkorrelation zwischen g -Faktor zu Z3 und Motivation bzw. Schulerfolg	194
13.8	Full-Forward Modell Fähigkeitsselbstkonzept Mathematik und Deutsch	196
13.9	Modellfit im MKSD – Fähigkeitsselbstkonzept Mathematik und Deutsch	199
13.10	Modelle zur Beurteilung eines Entwicklungstrends – FSK Mathematik	200
13.11	Full-Forward Modell intrinsische Motivation Mathematik und Deutsch	204
13.12	Modellfit im MKSD – Intrinsische Motivation Mathematik und Deutsch	205

13.13	Multigruppen Modelle Fähigkeitsselbstkonzept Mathematik mit Kontrolle von g	208
14.1	Intraklassenkorrelation der Variablen zu den verschiedenen Messzeitpunkten	214
14.2	Modell-Fit der latenten univariaten Modelle – Schulerfolg und g	217
14.3	Modell-Fit der latenten univariaten Modelle – Fähigkeitsselbstkonzept	218
14.4	Modell-Fit der latenten univariaten Modelle – Intrinsische Motivation	219
14.5	Parameterschätzung der best-fittenden latenten univariaten Modelle	220
14.6	Modell-Fit der latenten multivariaten Modelle	222
14.7	Bivariate Erblichkeit und bivariate geteilte und nichtgeteilte Umwelt	226
B.1	Gängige „Daumenregeln“ zur Bewertung der Modellgüte	297
B.2	Notation der Variablen im Pfadmodell der SEM	298
B.3	Gesamtstichprobe K_AS	299
B.4	Faktorladungen der Messmodelle in den Kohorten	300
B.5	Prüfung der Messinvarianz Zeit – Schulerfolg Mathematik und Deutsch	301
B.6	Prüfung der Messinvarianz Zeit – FSK Mathematik und Deutsch	302
B.7	Prüfung der Messinvarianz Zeit – IM Mathematik und Deutsch	303
B.8	Manifeste Interkorrelation zwischen Motivation und Schulnoten	304
B.9	Full-Forward Modell Fähigkeitsselbstkonzept Mathematik und Deutsch mit Kontrolle von g	305

Abkürzungsverzeichnis

<i>g</i> -Faktor	Generalfaktor der Intelligenz
r_A	Genetische Korrelation
r_C	Geteilte Umweltkorrelation
r_E	Nichtgeteilte Umweltkorrelation
r_{GE}	Gen-Umwelt Korrelation
GxU	Gen-Umwelt-Interaktion
AIC	Akaike Information Criterion
AM	Assortative Mating
AV	Abhängige Variable
BFLPE	Big-Fish-Little-Pond-Effekt
C.R.	Critical Ratio
CFI	Comparative Fit Index
CI	Condition Index
CLP	Cross-Lagged Panel
COTD	Children of Twins Design
CTCT	Cross-Trait Cross-Twin Correlation
CTD	Classical Twin Design
CU	Correlated Uniqueness
df	Degrees of Freedom
EEA	Equal Environments Assumption
EM	Expectation Maximazation
ETFD	Extended Twin Family Design
EZ	Eineiige Zwillinge
FSK	Fähigkeitsselbstkonzept
I/E-Modell	Internal/External-Frame-of-Reference-Modell
ICC	Intraklassenkorrelation
IM	Intrinsische Motivation
K_AS	Analysestichprobe
K_AS_2	Analysestichprobe der längsschnittlichen Betrachtung
KI	Konfidenzintervall
KoSMoS	Zwillingstudie zu Einflüssen von Kognitiven Fähigkeiten und selbst- eingeschätzter Motivation auf Schulerfolg
LOGIK	Longitudinalstudie zur Genese individueller Kompetenzen
LRT	Likelihood-Ratio-Test
MAR	Missing at random
MCAR	Missing completely at random

MHT	Moray House Test
MKSD	Multi-Kohorten-Sequenz-Design
ML-Schätzung	Maximum-Likelihood-Schätzung
MTFS	Minnesota Twin Family Study
MVA	Missing Value Analyse
NC	National Curriculum
NMAR	Not missing at random
NTFD	Nuclear Twin Family Design
OFSTED	Office for Standards in Education
PSQ	Physical Similarity Quotient
QTLs	Quantitative Trait Loci
REM	Reciprocal Effects Model
RMSEA	Root-Mean-Square-Error of Approximation
RNI	Relative Noncentrality Index
SCHOLASTIK	Schulisch organisierte Lernangebote und die Sozialisation von Talenten, Interessen und Kompetenzen
SE	Schulerfolg
SEM	Structural Equation Modeling
SES	sozioökonomischer Status
SRMR	Standardized Root Mean Square Residual
TEDS	Twins' Early Development Study
TIMSS	Third International Mathematics and Science Study
TLI	Tucker-Lewis-Index
TRd	Satorra-Bentler Chi-Quadrat Differenztest
TW	Toleranz-Wert
UK	Vereinigtes Königreich
UV	Unabhängige Variable
VIF	Variance Inflation Factor
Z1	Erster Erhebungszeitpunkt, 2005
Z2	Zweiter Erhebungszeitpunkt, 2007
Z3	Dritter Erhebungszeitpunkt, 2009
ZZ	Zweieiige Zwillinge

I. EINLEITUNG

Einleitung

Wohl kaum jemand wird der Schule, oder allgemeiner der Bildung, ihre hohe Relevanz in einer immer stärker an Leistung und Erfolg orientierten Gesellschaft absprechen. Bereits im Grundschulalter werden in der Wahrnehmung von Eltern und Schülern die Weichen für eine erfolgreiche Schullaufbahn gestellt, die wiederum als Wegbereiter für Erfolg im Erwachsenenalter gilt (Jimerson, Egeland & Teo, 1999; Sirsch, 2000). Schulnoten als eine Form der individuellen Leistungsevaluation bestimmen dabei zunächst über die Chancen einer Gymnasialempfehlung und im weiteren Verlauf über die Möglichkeit der Partizipation an einem höheren Bildungsweg. Speziell der Übergang von der Grund- in die Sekundarsschule wird hierbei in der Literatur als kritisches Lebensereignis (Filipp, 1995) bzw. als Entwicklungsaufgabe (Havighurst, 1972) beschrieben.

Nicht zuletzt aufgrund der alarmierenden Ergebnisse aus internationalen Schulleistungsstudien der letzten zehn Jahre werden die Bedingungsfaktoren der Schulleistungsentwicklung auch im Hinblick auf Fördermöglichkeiten und Veränderungen der schulischen Lernumwelt wieder rege diskutiert (Helmke & Schrader, 2006; Finsterwald, 2006). Die hohe Bedeutung, die der Bildung zugemessen wird, spiegelt sich dabei auch in den aktuellen politischen Diskussionen um die Bemühungen der Verbesserung und Weiterentwicklung des Bildungssystems wider, die alle das Ziel verfolgen, den Schülern eine Lernumgebung bereitzustellen, die es ihnen ermöglicht, ihr Leistungspotenzial optimal ausschöpfen zu können (siehe dazu z. B. die Ergebnisse des „Bildungsmonitor 2012“¹ oder die Zieldefinition des Ministeriums für Bildung und Forschung²).

Die Suche nach Einflussfaktoren auf die schulische Lern- und Leistungsentwicklung stellt entsprechend eine der Kernaufgaben der empirischen Bildungsforschung dar. In diesem Kontext ist davon auszugehen, dass sowohl Merkmale des Kindes und der Schule als auch familiäre Kontextvariablen zu berücksichtigen sind (Baumert, Becker, Neumann & Nikolova, 2009; Weinert, 2002), denn im Regelfall handelt es sich um vielschichtige funktionale Zusammenhänge und multiple Interaktionen zwischen diesen Determinanten, die sich in einem komplexen Gefüge gegenseitig beeinflussen (Helmke & Schrader, 2006). Darüber hinaus kann von reziproken Effekten zwischen den Determi-

¹ <http://www.insm-bildungsmonitor.de>

² <http://www.bmbf.de/de/90.php>

nanten und der Schulleistung in dem Sinne ausgegangen werden, dass das Vorliegen von Leistungen sowie Leistungsfortschritten und deren Wahrnehmung auf viele kognitive, motivationale und soziale Einflussfaktoren zurückwirken (Helmke & Weinert, 1997).

Eine breite empirische Basis identifiziert auf Seiten der individuellen Einflussfaktoren die kognitiven Eingangsvoraussetzungen der Kinder mit Korrelationen von ca. 0.50 als wichtigsten Prädiktor schulischen Erfolgs (Gustafsson & Undheim, 1996; Neisser et al., 1996). In anderen Worten bedeutet dies einerseits, dass *beachtliche* 25% der Varianz in schulischen Leistungen durch die Intelligenz der Kinder erklärt werden können. Es bedeutet gleichsam aber auch, dass *nur* 25% der Varianz durch Unterschiede in der kognitiven Leistungsfähigkeit erklärt werden und damit viel Raum für weichere, leichter beeinflussbare Merkmale wie Motivation, Attribution oder familiäre und schulische Kontextvariablen bleibt.

Im Hinblick auf relevante Charakteristika der Kinder gelten motivationale Variablen in der aktuellen Literatur als besonders einflussreiche Schülermerkmale die mit dem Leistungsverhalten in Beziehung stehen (Lüdtke, Köller, Artelt, Stanat & Baumert, 2002). Das gesteigerte Interesse an solchen motivationalen Merkmalen lässt sich auch damit erklären, dass diese als durch äußere Umwelteinflüsse relativ leicht modifizierbar angenommen werden (J. Heckhausen & H. Heckhausen, 2006).

Aus diesem Grund werden auch in der vorliegenden Studie motivationale Merkmale der Kinder, genauer das fachspezifische Fähigkeitsselbstkonzept und die intrinsische Motivation, hinsichtlich deren prädiktiver Validität neben allgemeinen kognitiven Fähigkeiten analysiert. In Ergänzung des bisherigen Forschungsstandes werden dabei im Rahmen eines Multi-Kohorten-Designs auch die längsschnittlichen Wirkzusammenhänge und hier speziell der Übergang zur weiterführenden Schule, beleuchtet.

Die Arbeit ist innerhalb der Saarbrücker *Zwillingsstudie zu Einflüssen von kognitiven Fähigkeiten und selbst eingeschätzter Motivation auf Schulerfolg* (KoSMoS; F. M. Spinath & Wolf, 2006) angesiedelt. Diese Längsschnittstudie wurde im Jahr 2005 initiiert und erfasste ein umfangreiches Set an individuellen Merkmalen und familiären Kontextvariablen die insgesamt dreimal im Abstand von jeweils zwei Jahren erhoben wurden. Zum ersten Messzeitpunkt nahmen insgesamt 408 Zwillingskinder im Alter zwischen 7 und 11 Jahren an der Untersuchung teil, zum dritten Messzeitpunkt in Jahr 2009 waren dies noch 209 Paare. Durch die Einbettung in dieses genetisch sensitive Design können die phänotypischen Ergebnisse um ätiologische Analysen erweitert werden, denn gerade wenn es um individuelle Determinanten schulischer Leistung geht, kann die Verhaltensgenetik wertvolle neue Erkenntnisse liefern. Erst durch die Untersuchung dieser Faktoren im genetisch sensitiven Design wird es möglich, den Ursachen interindividueller Unterschiede in diesen Variablen auf den Grund zu gehen und das diesbezüglich vorhandene Wissen zu erweitern (Deary, 2001). Zudem können unter Rückgriff auf

verhaltensgenetische Analysen die Ursachen von Stabilität und Veränderung des schulischen Erfolgs sowie dessen Prädiktoren einer eingehenderen Betrachtung unterzogen werden.

Verhaltensgenetische Forschung hat in diesem Zusammenhang den klaren Nachweis erbracht, dass interindividuelle Unterschiede in kognitiven Variablen zu einem großen Anteil genetischen Einflüssen unterliegen und dass deren Einfluss mit steigendem Lebensalter zunimmt (z. B. T. Bouchard & McGue, 1981; Deary, 2001; Plomin & F. M. Spinath, 2004). Auch für schulischen Erfolg liegen bereits einige Studien vor, die ebenfalls eine genetische Beeinflussung nahelegen (Johnson, McGue & Iacono, 2007; Kovas, Haworth, Dale & Plomin, 2007; Walker, Petrill, Spinath & Plomin, 2004). Die Ursachen interindividueller Unterschiede in motivationalen Variablen sind hingegen noch weitgehend ungeklärt und auch die Ursachen von Stabilität und Veränderung in diesen Variablen wurden bisher noch nicht untersucht.

Die vorliegende Studie liefert, vor allem durch die methodische Umsetzung innerhalb eines längsschnittlichen Multi-Kohorten-Designs über einen Zeitraum von sechs Jahren, einen wichtigen Beitrag zum Verständnis der Beziehung zwischen Motivation und schulischem Erfolg und ergänzt so die bereits vorhandene empirische Datenlage sowohl auf phänotypischer als auch ätiologischer Ebene. Innerhalb des Untersuchungsdesigns wird es möglich, auch die Bedeutung von Kohorten- und Entwicklungseffekten zu prüfen. Da die Kinder innerhalb des Untersuchungszeitraumes alle von der Grundschule in die weiterführende Schule wechseln und dieser Wechsel mit großen Veränderungen in der Schulumwelt verbunden ist (Sirsch, 2000), liegt die Annahme solcher Effekte nahe. Werden phänotypisch Entwicklungseffekte gefunden, können diese in einem nächsten Schritt durch gezielte ätiologische Analysen auf eine genetische bzw. umweltbedingte Beeinflussung hin überprüft werden. Weiterhin können durch die Ergänzung verhaltensgenetischer Analysen die Ursachen von Stabilität und Veränderung in den betrachteten Variablen aufgezeigt werden. Diesbezüglich wird durch die gewählte latente Modellierung der Variablen und die Berücksichtigung von Messfehlern die Bewertung der Stabilität zuverlässiger als in der Mehrzahl bereits existierender Studien (vgl. hierzu Kapitel 3.5 sowie Kapitel 9).

Zudem können die bisher meist aus dem amerikanischen Raum stammenden Studien durch eine umfangreiche deutsche Stichprobe (mit unterschiedlichem Schulsystem) ergänzt werden und schließlich wird es durch die Kontrolle allgemeiner kognitiver Fähigkeiten möglich, die von Intelligenz unabhängigen Beiträge zu bewerten. Dabei wird im Unterschied zur Mehrzahl vorhandener Studien eine Operationalisierung gewählt, die eine fachspezifische Bewertung der Zusammenhänge ermöglicht.

Die Arbeit gliedert sich in insgesamt fünf Teile: In einem ersten Teil werden zunächst

die zugrunde liegenden theoretischen Annahmen erläutert und ein zusammenfassender Überblick des empirischen Forschungsstandes zur Vorhersage schulischen Erfolgs durch allgemeine kognitive Fähigkeiten sowie Motivation gegeben. Eine Erläuterung der verhaltensgenetischen Methode ist ebenfalls Gegenstand dieser Darstellung.

Hieraus abgeleitet erfolgt im zweiten Teil die Herleitung der Fragestellungen und Hypothesen vorliegender Arbeit. In einem dritten Teil werden die KoSMoS Studie, das Vorgehen bei der Auswahl der Untersuchungsstichprobe sowie die eingesetzten Instrumente und Analyseverfahren, insbesondere die Multi-Kohorten-Sequenz-Analyse, beschrieben.

Der vierte Teil umfasst schließlich die Ergebnisse der durchgeführten phänotypischen und ätiologischen Analysen.

Die Arbeit schließt im fünften Teil mit einer zusammenfassenden Interpretation der Ergebnisse und den hieraus ableitbaren Implikationen sowie einer kritischen Darstellung der Grenzen der vorliegenden Arbeit. Zuletzt wird ein Ausblick auf weitere Forschungsfragen gegeben.

II. THEORETISCHE GRUNDLAGEN UND EMPIRISCHE BEFUNDE

Im theoretischen Teil der vorliegenden Arbeit wird zunächst das Kriterium Schulerfolg sowie Möglichkeiten zu dessen Operationalisierung betrachtet. Hinsichtlich der Prädiktoren allgemeine kognitive Fähigkeiten sowie Motivation werden die der Arbeit zugrunde gelegten theoretischen Modellvorstellungen erörtert und der empirische Forschungsstand zur prädiktiven Validität dieser Variablen dargestellt. Hierbei liegt der Fokus auf Studien und Modellen, die eine längsschnittliche Analyse der Zusammenhänge umsetzen. Abschließend wird eine Einführung in die Grundlagen der Verhaltensgenetik gegeben und der bisherige Forschungsstand zur Ätiologie von schulischem Erfolg, allgemeinen kognitiven Fähigkeiten und Motivation zusammenfassend dargestellt.

1 Das Kriterium Schulerfolg

Wie einleitend dargelegt, werden bereits im Grundschulalter die Weichen für den späteren schulischen Erfolg und damit in der Folge auch für den beruflichen Erfolg im Erwachsenenalter gestellt (Jimerson et al., 1999). Bei der Suche nach ausschlaggebenden Bedingungsfaktoren schulischen Erfolgs ist es in einem ersten Schritt notwendig, zwischen *Schulleistung* als Kompetenz in unterschiedlichen Aufgabenfeldern und *Schulerfolg* im Sinne fachspezifischer Noten zu unterscheiden.

Im Rahmen der vorliegenden Arbeit wird das Kriterium – der individuelle Schulerfolg – durch fachspezifische Noten operationalisiert. Aus diesem Grund wird im Folgenden zunächst eine definitorische Abgrenzung der beiden häufig im Kontext pädagogisch psychologischer Forschung verwendeten Begriffe *Schulleistung* und *Schulerfolg* vorgenommen. In einem zweiten Teil wird der Forschungsstand in Bezug auf die Entwicklung des individuellen Schulerfolgs über die Schulzeit hinweg dargestellt.

1.1 Operationalisierung

Unter *Schulleistungen* werden in der Literatur unterschiedliche Aspekte subsumiert, die von schulischem Leistungsverhalten, dem erreichten Bildungsniveau bis hin zu bereichsspezifischem Wissen reichen (Sauer & Gamsjäger, 1996; Helmke, 1997). Der Begriff bezieht sich dementsprechend sowohl auf Leistungen der Schule als auch der Schüler, wobei beide Aspekte bei der Entwicklung von Leistungsunterschieden zwischen Schülern beteiligt sind (Weinert, 2002). Als Indikatorvariablen dienen in der Regel Lehrerurteile, Ergebnisse in standardisierten Schulleistungstests oder die fachspezifischen Noten (Helmke & Schrader, 2006). Testdaten werden dabei auf Grund ihrer höheren Reliabilität, Objektivität und Validität meist als adäquateres Mittel zur Operationalisierung von Schulleistungen angesehen als fachspezifische Noten (vgl. z. B. Ingenkamp, 1971; Jensen, 1993). Schulnoten dienen jedoch im schulischen Alltag der Bestimmung von Schulleistungsunterschieden und können folglich zur Quantifizierung von interindividuellen Unterschieden und intraindividuellen Veränderungen herangezogen werden (Tent, 2006).

Die häufig angeführte Kritik an der Verwendung von Schulnoten als Kriterium bezieht sich überwiegend auf Merkmale, die in der Unterrichtspraxis kaum vermeidbar sind, wie z. B. Beobachtungsmängel, Erinnerungsfehler oder Urteilstendenzen (Tent, 2006). Noten beziehen sich zudem auf die relative Position des Schülers innerhalb des Klassenverbandes und werden durch Faktoren beeinflusst, die für die schulische Leistungsfähigkeit wenig relevant sind, so z. B. das Verhalten des Schülers oder seine sozialen Kompetenzen (Tent, 2006; Steinmayr & B. Spinath, 2009). Gleichwohl finden sich auch Belege für die Angemessenheit von Schulnoten zur Beurteilung von Schulleistungen. Beispielsweise klären mathematische Schulleistungstests nur geringfügig mehr Varianz in der Vorhersage der späteren Mathematiknote auf als die Vorjahresnote (Helmke, 1992).

Fachspezifische Noten stellen dementsprechend eine valide und ökonomische Alternative zu Schulleistungstests dar, insbesondere dann, wenn auf eine Stichprobe zurückgegriffen werden muss, die sich über ein weites Gebiet verteilt und damit, wie in der vorliegenden Studie, eine Face-to-Face Testung nicht möglich ist. Wenngleich objektive Tests eine direktere Erfassung der potenziellen *Schulleistung* ermöglichen, spiegeln Schulnoten den unmittelbaren *Schulerfolg* in adäquaterer Weise wider (Steinmayr & B. Spinath, 2008; für einen Vergleich von Schulleistungstest und Schulnoten siehe Willingham, Pollack & Lewis, 2002). Aus diesen Gründen wurde auch in KoSMoS auf fachspezifische Noten zur Operationalisierung des Schulerfolgs zurückgegriffen. Nachfolgend wird zur Schaffung begrifflicher Klarheit bei Rückgriff auf fachspezifische Noten der Begriff *Schulerfolg* und bei Rückgriff auf Daten aus Schulleistungstests der Begriff *Schulleistung* verwendet.

1.2 Entwicklung schulischen Erfolgs

Im Kontext der Forschung zu schulischem Erfolg und dessen Bedingungsfaktoren ist auch die Frage nach der Entwicklung schulischer Leistung im Sinne intra¹- und interindividueller² Stabilität über die Schuljahre hinweg von Relevanz.

Für die Grundschulzeit liegen in erster Linie Erkenntnisse in Bezug auf die Entwicklung verschiedener schulischer Leistungen im Sinne von Rechtschreib-, Lese- und Rechenkompetenzen vor. Diese sprechen einerseits für einen signifikanten Leistungszuwachs sowohl der Lese- und Rechtschreib- (z. B. Schneider, Stefanek & Dotzler, 1997) als auch der Rechenkompetenz (z. B. Reusser, 1997) bis zur vierten Grundschulklasse. Anderer-

¹ *Intraindividuelle Stabilität* oder *Niveau-Stabilität* liegt vor, wenn ein Individuum zu zwei Messzeitpunkten die gleiche Merkmalsausprägung aufweist (Schmitt, 2003).

² *Interindividuelle Stabilität* oder *relative Stabilität* liegt vor, wenn Merkmalsunterschiede zwischen Personen über die Zeit gleich bleiben (Schmitt, 2003).

seits erweisen sich die interindividuellen Unterschiede in den schulischen Leistungen als vergleichsweise entwicklungsstabil, d. h. Rangunterschiede zwischen Schülern bleiben in der Regel erhalten (Ditton, 2010a; Grube & Hasselhorn, 2006; Helmke & Schrader, 2007). Gleiches kann auch für gemittelte Schulnoten in der Grundschule angenommen werden. Im Verlauf der Grundschule verschlechtern sich die Schulnoten zwar kontinuierlich, ihre insgesamt eher rechtssteile und damit von der Normalverteilungsannahme abweichende Verteilung, bleibt aber insgesamt bestehen (Helmke, 1997; Tent, 2006). Zudem erweisen sich Schulnoten interindividuell als außerordentlich stabil: Über einen Zeitraum von drei Jahren werden Stabilitäten von ca. 0.80 berichtet (Böhnel, 1996; Tent, 2006).

Neben dieser singulären Betrachtung der Grundschulzeit liefern insbesondere Längsschnittstudien zur Entwicklung schulischen Erfolgs beim Wechsel von der Grund- in die Sekundarschule aufschlussreiche Ergebnisse. Das breite Interesse an der Prognose schulischen Erfolgs nach dem Übergang in das weiterführende Schulsystem lässt sich sicherlich auch durch die entscheidende Rolle, die den Zensuren am Ende der vierten Klasse für die letztliche Schulempfehlung zukommt, erklären. Die Entscheidung über die Art der weiterführenden Schule und damit über den angestrebten Bildungsabschluss, stellt die erste Weiche für die spätere berufliche Perspektive der Kinder und wird von Eltern und Schülern häufig als wichtige Statusvoraussetzung wahrgenommen (Sirsch, 2000).

In der Literatur wird der Wechsel von der Grundschule in die weiterführende Schule häufig unter Rückgriff auf Havighursts' Theorie der Entwicklungsaufgaben (Havighurst, 1972) beleuchtet (z. B. Ball, Lohaus & Miebach, 2006; Eccles & Midgley, 1989; Philipp, 1995; Koch, 2004). Diese werden als universelle, an das Lebensalter gebundene Anforderungen verstanden, mit denen sich jedes Individuum im Laufe seines Lebens konfrontiert sieht (Schmid, 1998). Der Wechsel in das weiterführende Schulsystem fällt in die vierte von neun Entwicklungsperioden (*Mittleres Schulalter*; Havighurst, 1972) und erfordert von den Kindern eine Anpassung an neue soziale sowie leistungsbezogene Anforderungen (Büchner & Koch, 2001; Hacker, 1988; Koch, 2004). Aus entwicklungspsychologischer Sicht fällt der Zeitpunkt des Wechsels, der für deutsche Kinder etwa im Alter von zehn Jahren stattfindet, dabei in eine „relativ einheitliche, undramatische, mehr oder minder kontinuierliche Wachstums- und Bildungsperiode körperlicher, geistiger wie persönlicher Merkmale“ (Weinert & Stefanek, 1997, S. 439). Dennoch kann die Anpassung an diese neuen Gegebenheiten in Abhängigkeit von z. B. der Reife der Kinder, ihren Copingressourcen und familiären Faktoren interindividuell unterschiedlich erfolgreich verlaufen (Crockett, Petersen, Graber, Schulenberg & Ebata, 1989; Duchesne, Larose & Guay, 2005).

Aus empirischer Sicht bestätigt sich der negative Trend der Entwicklung schulischen Erfolgs auch nach dem Übergang in die weiterführende Schule. Die große Mehrzahl an

Studien belegt eine Verschlechterung der Noten in der Sekundarstufe nach jedem Schuljahr (z. B. Alspaugh, 1998; Ball et al., 2006; Seidman, Allen, Aber, Mitchell & Feinman, 1994), wobei die Verschlechterung im Jahr unmittelbar nach dem Wechsel besonders gravierend auszufallen scheint (Blyth, Simmons & Carlton-Ford, 1983; Crockett et al., 1989). Zudem scheint es einen schultypenspezifischen Trend der Schulleistungsentwicklung dahingehend zu geben, dass vor allem Gymnasialschüler eine Verschlechterung in ihren Schulnoten aufweisen, während Hauptschüler einen Anstieg in ihren Noten verzeichnen (Ball et al., 2006; Tent, 2006). Ball et al. (2006) erklären diesen Befund mit einer Veränderung der leistungsbezogenen Anforderungen und der Bildung homogener Lerngruppen.

Eine häufig gestellte Frage in diesem Kontext betrifft letztlich die Wichtigkeit des Alters beim Übergang in die Sekundarstufe. Hier ist zunächst die Generalisierbarkeit der meist aus dem amerikanischen Raum vorliegenden Forschungsbefunde kritisch zu hinterfragen. Zum einen finden sich generelle Unterschiede in den Schulsystemen zwischen Deutschland und den USA dahingehend, dass amerikanische Schüler bei Abschluss der „Elementary School“ in der Regel ein bis zwei Jahre älter sind als deutsche Schüler. Zum anderen fallen Brüche im Sozialgefüge der Kinder in den USA durch den gemeinsamen Wechsel eventuell weniger gravierend aus und sind mit einer weniger starken Leistungsdifferenzierung verbunden (Blyth et al., 1983; Finsterwald, 2006). Die Befunde verschiedener Studien sprechen allerdings eher für einen negativen Effekt der Häufigkeit des Schulwechsels auf die akademischen Leistungen und weniger für einen Effekt des Alters per se (z. B. Alspaugh, 1998; Felner, Primavera & Cauce, 1981; Schulenberg, Asp & Petersen, 1984; Seidman et al., 1994; Zimmerman, Copeland, Shope & Dielman, 1997).

Letztlich bleibt festzuhalten, dass der Übergang in das weiterführende Schulsystem und die damit verbundenen Herausforderungen von Kindern und Eltern gleichermaßen als wichtiger biografischer Einschnitt wahrgenommen werden, der nicht nur mit positiven Erwartungen sondern auch mit Befürchtungen und Ängsten verbunden ist (Koch, 2004; Sirsch, 2000; Wiederhold, 1991). Neben einer Verschlechterung der Schulnoten werden häufig auch Veränderungen des Selbstkonzeptes, Schulangst und ein Wandel in der Motivation berichtet (Sirsch, 2000). Auch vor diesem Hintergrund rückt die Frage nach potenziellen Einflussmöglichkeiten auf die Schulleistung in den Mittelpunkt des Interesses. Die Identifikation solcher Faktoren – sowohl auf individueller als auch familiärer Ebene – eröffnet Möglichkeiten der Intervention und gezielten Förderung in Bezug auf bestimmte Kompetenzen. Als familiäre Einflussfaktoren werden häufig der Erziehungsstil oder das Engagement der Eltern hinsichtlich schulischer Belange untersucht. Auf Seiten individueller Determinanten, denen ein besonderes Gewicht bei der Vorhersage schulischen Erfolgs zugesprochen wird (Helmke, 1997), finden sich zunächst grundlegende biologische Merkmale (Alter, Geschlecht, Erkrankungen etc.),

Persönlichkeitseigenschaften (wie z. B. das individuelle Temperament) sowie affektive Grundzüge des Kindes. Ferner sind kognitive Fähigkeiten sowie motivationale Einflussfaktoren den individuellen Determinanten zuzurechnen (Helmke & Schrader, 2006).

Die beiden letztgenannten Komponenten – kognitive Fähigkeiten und bestimmte motivationale Voraussetzungen – stehen im Mittelpunkt der vorliegenden Arbeit. Nach einer theoretischen Einführung beider Konzepte wird jeweils der bisherige Forschungsstand zur Bedeutung dieser Faktoren für akademische Leistungen einerseits und deren Stabilität über die Schullaufbahn hinweg andererseits zusammenfassend dargestellt. Aufgrund der inhaltlichen Breite dieser Konstrukte und einer kaum mehr überschaubaren Fülle empirischer Studien kann dabei kein Anspruch auf Vollständigkeit erhoben werden.

2 Bedingungsfaktoren schulischen Erfolgs: Allgemeine kognitive Fähigkeit

„Wissen schlägt Intelligenz“ – so der Titel eines Interviews mit der Lernforscherin Elsbeth Stern in der *Zeit online* vom 26. Juni 2003. Wenngleich insbesondere Ergebnisse aus der Expertiseforschung belegten, dass eine weniger günstige geistige Voraussetzung kompensiert werden könne, bleibe Intelligenz „ein sehr sinnvolles wissenschaftliches Konstrukt, um die unterschiedliche geistige Leistungsfähigkeit von Menschen zu erklären, die die gleichen Lerngelegenheiten hatten. Intelligenz beeinflusst ohne Zweifel den Erfolg in der Schule und im Beruf“ (Stern, 2003). Die Kombination aus Wissen und Intelligenz sei die beste Voraussetzung, gute Leistungen beispielsweise im Fach Mathematik zu erzielen. Der große Vorteil intelligenter Kinder liege vor allem darin, dass sie leichter und schneller lernen. Diese Einschätzung teilen auch Helmke und Schrader (2006), die speziell für den schulischen Bereich annehmen, dass intelligenterer Schüler schneller und effektiver lösungsrelevante Regeln erkennen und Probleme lösen, was wiederum den Aufbau einer besser vernetzten, flexibler nutzbaren und „intelligenter“ organisierten Wissensstruktur erleichtert (Helmke & Schrader, 2006, vgl. S.84). Ob sich dieser Vorteil letztlich auch in einer besseren Schulleistung widerspiegelt, wurde in der pädagogisch-psychologischen Forschung vielfach diskutiert (Helmke, 1997; siehe hierzu Abschnitt 2.3).

Durch die große Beachtung, die der Begriff *Intelligenz* sowohl in der wissenschaftlichen als auch der gesellschaftlichen Diskussion erfahren hat, entwickelte sich eine Vielzahl an Definitionsversuchen und Vorstellungen über deren Struktur (Gustafsson & Undheim, 1996; Sternberg, 2004), die im Rahmen dieser Arbeit jedoch nicht erschöpfend dargestellt werden sollen. Es wird hingegen zunächst eine allgemeine Begriffsbestimmung des Konstruktes *Intelligenz* vorgenommen, um für die weiteren Ausführungen eine gemeinsame Basis zu gewährleisten. Anschließend werden die Generalfaktorentheorie (Spearman, 1904) und das Three Stratum Modell (Carroll, 1993) als theoretischer Rahmen der vorliegenden Arbeit vorgestellt. Abschließend werden empirische Befunde zur prognostischen Relevanz kognitiver Fähigkeiten für schulischen Erfolg, sowie deren Entwicklungsstabilität zusammenfassend dargestellt.

2.1 Begriffsbestimmung

Intelligenz gilt nicht nur als eines der meist untersuchten psychologischen Konstrukte, sondern ist auch dasjenige mit der weitreichendsten prädiktiven Validität für verschiedenste Bereiche, z. B. den Ausbildungs- und Studienerfolg oder die Lebenserwartung (Amelang, Bartussek, Stemmler & Hagemann, 2006; Whalley & Deary, 2001).

Historisch gesehen lassen sich insbesondere zwei Meilensteine identifizieren, welche die Forschung in diesem Bereich maßgeblich geprägt haben und diese bis heute beeinflussen. Im Jahr 1904 veröffentlichte Charles Spearman im *American Journal of Psychology* eine der ersten theoretischen Arbeiten zur allgemeinen Intelligenz (vgl. hierzu Abschnitt 2.2.1), mit der er den Grundstein der Intelligenzstrukturforschung legte (*General Intelligence objectively determined and measured*; Spearman, 1904). Diese hat sich in erster Linie der Frage gewidmet, wie viele und welche Fähigkeitsbereiche die allgemeine kognitive Fähigkeit ausmachen und wie deren struktureller Aufbau anzunehmen sei. Weiterhin können für den Bereich der psychometrischen Erfassung von Intelligenz die Arbeiten von Binet und Simon, deren Intelligenztest im Jahr 1905 erschien, als wegweisend angesehen werden (Binet & Simon, 1905, 1904a, 1904b).

Intelligenz stellt dementsprechend ein komplexes, hypothetisches Konstrukt dar, welches als solches nicht direkt beobachtbar ist, sondern vielmehr aus dem Verhalten von Personen abgeleitet werden muss (Holling, Preckel & Vock, 2004). Mit diesem Umstand ging im Verlauf des letzten Jahrhunderts eine heftige Kontroverse über die grundlegende Struktur und die Definition der Intelligenz einher, die neben einer Vielzahl an Modellvorschlägen auch zu zahlreichen verbalen Definitionsversuchen geführt hat. Die Schwierigkeit, eine allgemein anerkannte Definition von Intelligenz zu finden, an der es im Übrigen bis heute fehlt (Funke & Vaterrodt-Plünnecke, 2004; Gustafsson & Undheim, 1996), wird bei Betrachtung verschiedener Definitionen ersichtlich. So wird Intelligenz beispielsweise als „zusammengesetzte oder globale Fähigkeit des Individuums, zweckvoll zu handeln, vernünftig zu denken und sich mit seiner Umgebung wirkungsvoll auseinander zu setzen“ beschrieben (Wechsler, 1944, S. 19). Weiter skizziert Robert J. Sternberg in seiner Konzeption der sog. „successful intelligence“ diese als das Ergebnis verschiedener basaler (z. B. Anpassung, Veränderung und Wahl von Umwelten) und mentaler (analytische, kreative und praktische Fähigkeiten) Prozesse, die sich je nach Kontext unterschiedlich manifestieren (vgl. z. B. Sternberg, 2003, 2005). Intelligenz lässt sich demnach durch eine Vielzahl abstrakter kognitiver Teilfähigkeiten charakterisieren, von denen nur einige psychometrisch erfassbar sind (Brocke & Beau-ducel, 2001; Jensen, 1998). Nichtsdestotrotz kann festgehalten werden, dass innerhalb der psychologischen Forschung sehr wohl eine klare und konsensfähige Vorstellung darüber existiert, welche spezifischen Fähigkeiten Intelligenz umfasst, wie bedeutsam die

Vorhersagekraft dieses Konstruktes ist und mit welcher Art von Aufgaben sie am besten erfasst werden kann (z. B. Carroll, 1993; Gottfredson, 2002; Klauer, 2006; Plomin & F. M. Spinath, 2004; Sternberg & Detterman, 1986). So wird davon ausgegangen, dass sich *Intelligenz* durch ein hierarchisch organisiertes Set an spezifischen kognitiven Fähigkeiten auszeichnet. Diese Fähigkeiten umfassen u.a. die Bereiche Wortschatz, verbale Produktion, numerische Fähigkeiten, räumliche Fähigkeiten, Gedächtnisleistung, Wahrnehmungsgeschwindigkeit und schlussfolgerndes Denken.

Im Kontext der vorliegenden Arbeit wird Intelligenz als *allgemeine kognitive Fähigkeit* verstanden, die unter anderem Problemlösefähigkeit, Verständnis komplexer Sachverhalte, Planungsfähigkeit, Abstraktionsvermögen, schlussfolgerndes Denken, schnelle Auffassungsgabe, sowie die Fähigkeit zum erfahrungsbasierten Lernen umfasst (Gottfredson, 1997a). Als Synonym zur allgemeinen kognitiven Fähigkeit wird häufig von einem allgemeinen *g-Faktor* gesprochen, der die Gemeinsamkeiten dieser heterogenen Prozesse beinhaltet. Psychometrisch repräsentiert *g* den Varianzanteil, der allen unterschiedlichen Maßen gemeinsam ist (Jensen, 2002). Diese psychometrische Herangehensweise zielt auf die Erfassung und Erklärung individueller Differenzen innerhalb und zwischen Altersstufen ab (Gustafsson & Undheim, 1996; Hasselhorn & Grube, 1997) und ist somit auch für diese Arbeit grundlegend. Der Erkenntnisgewinn erfolgt hier vor allem durch die Analyse messbarer Leistungen in psychometrischen Intelligenztests. Leistungen in diesen Tests dienen wiederum als Grundlage verschiedener Strukturmodelle der Intelligenz, in denen die Fähigkeiten mittels faktorenanalytischer Methoden zu übergeordneten Faktoren zusammengefasst werden (Holling et al., 2004).

Im Folgenden werden zwei faktorenanalytisch begründete Modelle beschrieben, die sich aus der psychometrischen Tradition heraus entwickelt haben und die oben angenommene psychometrische Definition allgemeiner kognitiver Fähigkeit adäquat repräsentieren: die Generalfaktorentheorie von Charles Spearman (Spearman, 1904) und die Three-Stratum-Theorie von John B. Carroll (Carroll, 1993). Für eine umfassende Darstellung weiterer Strukturmodelle sei an dieser Stelle auf Amelang et al. (2006) oder Holling et al. (2004) verwiesen.

2.2 Strukturmodelle der allgemeinen kognitiven Fähigkeit

2.2.1 Die Generalfaktorentheorie von Charles Spearman

Im Jahr 1904 formulierte Charles Spearman mit seiner *Generalfaktorentheorie* oder auch *Zwei-Faktoren-Theorie* das erste Faktorenmodell der Intelligenz, welches bis heu-

te die Operationalisierung und die Art und Weise, in der Intelligenz getestet wird, nachhaltig beeinflusst (Spearman, 1904). Aus der Beobachtung heraus, dass Leistungen in unterschiedlichen kognitiven Tests positiv miteinander korrelierten, schloss Spearman, dass kognitiven Leistungen ein gemeinsamer Faktor zugrunde liegt. Diesen Faktor bezeichnete er als *Generalfaktor* (g ; g -Faktor). Zusätzlich nahm er für jeden Test spezifische Fähigkeiten an, die zusammen mit g die Testleistung erklären sollten, wobei die Lösung der einzelnen Aufgabentypen unterschiedlich hohe g -Anteile erforderte (Holling et al., 2004). Darüber hinaus nahm er an, dass der Zusammenhang zwischen verschiedenen Subtests ausschließlich durch ihre jeweilige Korrelation zu g bestimmt sei.

Der g -Faktor spiegelt demnach individuelle Leistungsunterschiede in Testaufgaben wider (Jensen, 2002) und ist somit im Gegensatz zur Bedeutung des Wortes *Intelligenz* fest umrissen. Spätere Studien konnten allerdings zeigen, dass ein Modell mit einem Generalfaktor und einer für jeden Subtest spezifischen Fähigkeit nicht ausreichend ist, um die empirischen Daten hinreichend zu beschreiben und stattdessen ein komplexeres Modell mit mehreren Hierarchieebenen als adäquater angenommen werden kann (Gustafsson & Undheim, 1996).

2.2.2 Das Three-Stratum-Modell von John B. Carroll

Ein aktuellerer Ansatz der weithin Anerkennung findet, wurde von John B. Carroll vorgelegt (Carroll, 1993). Sein *Three-Stratum-Modell* (vgl. Abbildung 2.1) wird von vielen Wissenschaftlern als angemessener Integrationsversuch unterschiedlicher Strukturmodelle, z. B. der Generalfaktorentheorie von Spearman (1904), der Primärfaktorentheorie nach Thurstone (1947), des Konzepts der fluiden und kristallinen Intelligenz nach Cattell (1963) und der hierarchischen Intelligenzmodelle von Vernon (1964), betrachtet.

Carrolls Modell basiert auf einer faktorenanalytischen Re-Analyse von insgesamt 461 Datensätzen aus der Intelligenzforschung, die mit sehr heterogenen Tests und Aufgabentypen gewonnen wurden. Hieraus entwickelte Carroll ein dreistufiges, hierarchisches Strukturmodell der Intelligenz mit g als allgemeinste Fähigkeit an der Spitze (Stratum III) und spezifischeren Fähigkeiten auf untergeordneten Hierarchieebenen (Carroll, 1993, 1997; Gustafsson & Undheim, 1996; Sternberg, 2000). Die allgemeine Intelligenz wird durch komplexe kognitive Prozesse höherer Ordnung bestimmt und liegt allen intellektuellen Aktivitäten zugrunde. In Carrolls Theorie ist die allgemeine Intelligenz acht Fähigkeiten mittlerer Generalität übergeordnet, die ihrerseits wiederum spezifische Fähigkeiten beeinflussen. Die Faktoren mittlerer Generalität werden auf einer zweiten Ebene angesiedelt (Stratum II) und umfassen die fluide und kristalline Intelligenz, die allgemeine Gedächtnisfähigkeit, die visuelle sowie die auditive Wahrnehmung,

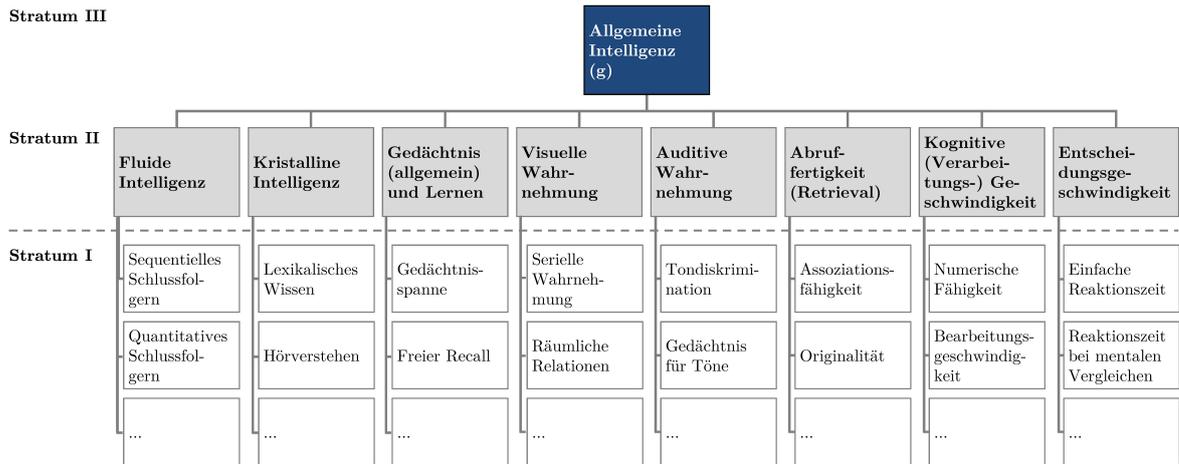


Abbildung 2.1. Das Three-Stratum Modell von Carroll. Eigene Darstellung in Anlehnung an Abbildung 15.1 aus Carroll, 1993, S. 626; Die spezifischen Fähigkeiten der untersten Ebene sind nur Beispiele von zahlreichen möglichen.

die Abruffähigkeit, die kognitive Verarbeitungsgeschwindigkeit und schließlich die Entscheidungsgeschwindigkeit. Zudem wird angenommen, dass der Zusammenhang mit g umso niedriger ausfällt, je weiter rechts in der Abbildung die Fähigkeit aufgeführt ist. Die fluide und kristalline Intelligenz hängen demnach am stärksten mit g zusammen. Diesen acht Fähigkeiten sind 69 relativ spezifische Fähigkeiten auf Stratum I zugeordnet, wobei die jeweilige Zuordnung nicht bedeutet, dass diese Fähigkeit nur von der ihr übergeordneten abhängt. Vielmehr werden Mischformen aus den Fähigkeiten von Stratum II angenommen. Intellektuelle Fähigkeiten werden demzufolge als komplex und miteinander in Beziehung stehend und als nicht unabhängig voneinander angenommen.

2.2.3 Abschließende Bewertung

Insgesamt liegt mit dem Three-Stratum-Modell ein umfassendes, theoretisch sowie empirisch fundiertes Gesamtmodell der Struktur allgemeiner kognitiver Fähigkeit vor. Die Konzeption des g -Faktors – der auch im Three-Stratum-Modell als generellste Fähigkeit Eingang gefunden hat – hat sich in der Intelligenzforschung und hier speziell in psychometrischen Ansätzen durchgesetzt:

„ g has become the working definition of intelligence for most researchers, because it is a stable, replicable phenomenon that – unlike the IQ score – is independent of the ‘vehicles’ (tests) for measuring it“

(Gottfredson, 2002, S. 27)

Im Mittel klärt der g -Faktor des Stratums III etwa 40% der gesamten Varianz untersuchter Leistungen auf (Carroll, 1993). Neben seiner faktorenanalytischen Bedeutung kann er inhaltlich als Generalfaktor verstanden werden, der mit unterschiedlicher Gewichtung Grundlage aller kognitiver Fähigkeiten oder Leistungen in verschiedenen Lebensbereichen darstellt (Gottfredson, 1997b). Die breite Akzeptanz des g -Faktors geht dementsprechend primär auf die Tatsache zurück, dass er mit verschiedenen externen Lern- und Leistungsvariablen höher korreliert als irgendein anderer Faktor oder eine Kombination mehrerer Faktoren und er eine hohe prädiktive Validität über verschiedene Tests und Situationen hinweg besitzt (Gottfredson, 2002; Humphreys & Stark, 2002). Weiterhin lässt sich die Existenz eines generellen Faktors kognitiver Fähigkeiten auch mit Blick auf empirische Studien klar belegen (vgl. z. B. Gottfredson, 1997b; Gustafsson & Undheim, 1996; Johnson, Nijenhuis & Bouchard, 2008). Ferner korrelieren die g -Faktoren aus unterschiedlichen Testbatterien untereinander sehr hoch ($r > 0.90$; Jensen, 1998). Darüber hinaus klären spezifischere Fähigkeiten (Stratum II) und hier insbesondere Gedächtnis, Verarbeitungskapazität sowie verbale und räumliche Fähigkeiten (Deary, 2001), verbleibende Varianz auf. Mit Blick auf Ergebnisse aus ätiologischen Studien kann schließlich auch auf genetischer Ebene vom Vorliegen eines g -Faktors ausgegangen werden (siehe Abschnitt 4.6.2; vgl. z. B. D. Luo, Petrill & Thompson, 1994; Plomin & F. M. Spinath, 2002).

Die hier vorgestellte Taxonomie liefert dementsprechend eine adäquate Möglichkeit zur Beschreibung interindividueller Differenzen in allgemeinen kognitiven Fähigkeiten. Es wird daher im Folgenden inhaltlich von einer hierarchischen Organisation der allgemeinen kognitiven Fähigkeit ausgegangen, die psychometrisch als g -Faktor aus den vorliegenden empirischen Daten extrahiert wird.

Nachfolgend wird auf Basis empirischer Befunde die Relevanz der allgemeinen kognitiven Fähigkeit zur Vorhersage des Schulerfolgs sowie deren individuelle Entwicklungsstabilität bewertet.

2.3 Vorhersage von Schulerfolg durch allgemeine kognitive Fähigkeiten

Befunde aus zahlreichen Studien bestätigen Intelligenz zweifelsfrei als validesten Einzelprädiktor akademischer Leistungen (vgl. z. B. Gottfredson, 2002; Heller, 1997; Helmke & Schrader, 2006; Kuncel, Hezlett & Ones, 2004; Sauer & Gamsjäger, 1996; B. Spinath, Spinath, Harlaar & Plomin, 2006). Die berichteten Korrelationen zwischen allgemei-

ner Intelligenz und Schulerfolg (operationalisiert durch fachspezifische Noten) liegen im Durchschnitt bei 0.50 (Cattell & Butcher, 1968; Gustafsson & Undheim, 1996; Helmke, 1997). Intelligenz klärt demnach im Schnitt 25% der Varianz im Merkmal Schulerfolg auf. Die Höhe dieses Zusammenhangs findet auch in einschlägigen Metaanalysen Bestätigung (Fraser, Walberg, Welch & Hattie, 1987; Gustafsson & Undheim, 1996; Roth, Becker, Spengler & Spinath, 2011; Neisser et al., 1996; Wang, Haertel & Walberg, 1993).

Trotz dieser einheitlichen Befundlage sollte nicht unerwähnt bleiben, dass sich in Abhängigkeit vom ausgewählten Testverfahren, der gewählten Operationalisierung des Kriteriums Schulleistung, sowie dem Schultyp, dem Alter und der Herkunft des Schülers erhebliche Schwankungen in den berichteten Zusammenhängen finden ($r = 0.10$ bis $r = 0.90$; vgl. Gustafsson & Undheim, 1996; Heller, 1997). So liegen beispielsweise bei der Verwendung standardisierter Leistungstests als Kriterium die beobachteten Korrelationen in einem höheren Bereich ($r = 0.60$ bis $r = 0.90$; Deary, Strand, Smith & Fernandes, 2007; M. C. Frey & Detterman, 2004; Rindermann, 2006; Steinmayr & B. Spinath, 2009). Diese hohen Korrelationen erklären sich durch eine gewisse Varianzüberlappung von Schulleistungstests und Intelligenztests und sind aus diesem Grund teilweise als methodisches Artefakt zu interpretieren (Duckworth, Quinn & Tsukayama, 2011; Steinmayr & B. Spinath, 2009). Werden hingegen Schulnoten als Kriterium herangezogen, fallen die Zusammenhänge mit dem erreichten Intelligenztestwert erwartungsgemäß geringer aus. Die höchsten Korrelationen zeigen sich hier üblicherweise für das Fach Mathematik ($r = 0.40$; vgl. Amthauer, Brocke, Liepmann & Beauducel, 2001).

Mit Blick auf den Schultyp, und hier insbesondere im Vergleich der Grundschule zur Sekundarschule, kann zunächst festgehalten werden, dass die verbale Intelligenz zumindest im Grundschulbereich eine bessere Prädiktionskraft aufzuweisen scheint, als die nonverbale Intelligenz (Gustafsson & Undheim, 1996). Weiterhin fallen die berichteten Korrelationen in der Grundschule höher aus als in der Sekundarschule ($r = 0.60$ bis $r = 0.70$ vs. $r = 0.40$ bis $r = 0.50$). Dieser Trend wird allerdings eher auf eine zunehmende Selektion und der damit einhergehenden Varianzeinschränkung, denn auf den Schultyp oder das Alter der Kinder zurückgeführt (Cattell & Butcher, 1968; Humphreys & Stark, 2002).

Schließlich finden sich häufig höhere Korrelationen zwischen spezifischen Fähigkeiten (z. B. verbalen Fähigkeiten) und denjenigen Schulfächern, in denen diese Fähigkeit besonders relevant ist (z. B. Deutsch; siehe Amthauer et al., 2001). Dieser Zusammenhang konnte allerdings nicht in allen Studien bestätigt werden (B. Spinath et al., 2006).

Die prognostische Validität von Intelligenztests für die spätere Schulleistung kann da-

mit als gut belegt gelten. Der g -Faktor liefert dementsprechend eine sehr ökonomische Möglichkeit, generellen akademischen Erfolg vorherzusagen (z. B. Gottfredson, 2002).

2.4 Individuelle Entwicklungsstabilität allgemeiner kognitiver Fähigkeiten

Neben der Bewertung der prognostischen Relevanz kognitiver Fähigkeiten ist auch die Frage nach deren Entwicklungsverlauf im Kontext dieser Arbeit von Interesse. Nachfolgend wird die Befundlage zur Frage der strukturellen Konstanz sowie der interindividuellen Stabilität kognitiver Fähigkeiten überblicksartig dargestellt, wobei die empirische Evidenz insgesamt sowohl für eine hohe strukturelle als auch interindividuelle Stabilität allgemeiner kognitiver Fähigkeiten spricht.

Die Frage nach dem Vorliegen struktureller Stabilität ist insbesondere vor dem Hintergrund einer intraindividuellen Veränderung kognitiver Fähigkeiten über die Lebensspanne von Interesse. Die individuelle Ausprägung in kognitiven Merkmalen unterliegt vor allem im frühen Kindesalter aufgrund rapider Entwicklungsvorgänge noch relativ starken intraindividuellen Schwankungen, stabilisiert sich jedoch mit steigendem Alter (Gustafsson & Undheim, 1996; Holling et al., 2004; Humphreys & Stark, 2002; Moffitt, Caspi, Harkness & Silva, 1993; Wilson, 1983). Ungeachtet dieser Schwankungen finden sich in verschiedenen empirischen Studien Belege für die Vergleichbarkeit der grundlegenden Struktur kognitiver Fähigkeiten in verschiedenen Altersbereichen, und damit auch für eine generelle Gültigkeit des zuvor beschriebenen Three-Stratum-Modells. Es konnte gezeigt werden, dass zwischen Kindern und Jugendlichen (Carroll, 1993; Kumbinger & Wurst, 2000) sowie zwischen Kindern und Erwachsenen (Jäger et al., 2004) keine Unterschiede hinsichtlich des strukturellen Aufbaus kognitiver Fähigkeiten vorliegen. Bisher noch nicht hinreichend beantwortet ist hingegen die Frage, ob es eine Verschiebung der Wichtigkeit einzelner Fähigkeiten bzw. eine inhaltliche Veränderung einzelner Faktoren im Entwicklungsverlauf gibt (Holling et al., 2004).

Neben der Bestätigung der strukturellen Stabilität lässt sich aus den Ergebnissen zahlreicher Studien ableiten, dass es sich auch in Bezug auf die interindividuelle Stabilität um einen äußerst stabilen Trait handelt (für einen Überblick der wichtigsten Studien vgl. Deary, Whalley, Lemmon, Crawford & Starr, 2000), für den sich bereits ab einem Alter von 10 Jahren relativ hohe Stabilitäten der interindividuellen Unterschiede finden (Gustafsson & Undheim, 1996). Insgesamt liegen die Stabilitätskoeffizienten für die meisten Intelligenztests in etwa bei $r_{tt} = 0.90$ (Amelang et al., 2006). Selbst bei längeren Zeitintervallen zwischen den Erhebungen ist noch von einer außerordentlich

hohen interindividuellen Stabilität auszugehen (zwischen $r_{tt} = 0.89$ und $r_{tt} = 0.96$ über einen Zeitraum von sieben Jahren; Schaie, 1996). Das bisher längste Wiederholungsintervall wurde in einer Stichprobe realisiert, die erstmals im Alter von 11 Jahren und erneut als 80-jährige mit dem Moray House Test (MHT; Scottish Council for Research in Education, 1933) untersucht wurden. Die berichtete Rangreihenstabilität von $r_{tt} = 0.66$ spricht auch über nahezu die gesamte Lebensspanne hinweg für eine hohe interindividuelle Stabilität allgemeiner kognitiver Fähigkeiten (Deary, Whiteman, Starr, Whalley & Fox, 2004). Hierbei weisen Fähigkeiten mit größerer Nähe zur kristallinen Intelligenz (z. B. Alltagswissen) eine höhere Stabilität auf als solche, die eher der fluiden Intelligenz (z. B. schlussfolgerndes Denken) zuzurechnen sind ($r_{tt} = 0.66$ vs. $r_{tt} = 0.45$, Deary et al., 2004; $r_{tt} = 0.80$ vs. $r_{tt} = 0.60$, Schwartzman, Gold, Andres, Arbuckle & Chaikelson, 1987).

Mit Blick auf das Kindesalter wird die Befundlage erwartungsgemäß heterogener. Mit niedrigerem Lebensalter der Probanden sinkt auch die Stabilität der allgemeinen kognitiven Fähigkeit. Eine naheliegende Erklärung für diese geringere Stabilität liegt in den Durchführungsbedingungen von Leistungstests. Diese sind für diesen Altersbereich weniger stringent zu standardisieren, zudem ist eine angemessene Motivation der Kinder schwieriger zu erreichen und aufrechtzuerhalten (Amelang et al., 2006). Hinzu kommen die bereits angesprochenen raschen Entwicklungsvorgänge der frühen Kindheit, die tiefgreifende Veränderungen bewirken. Diese Bedingungen könnten insgesamt dazu führen, dass zu unterschiedlichen Messzeitpunkten auch unterschiedliche Fähigkeiten gemessen wurden (Holling et al., 2004). Dennoch berichten verschiedene Studien bereits im jungen Kindesalter beachtlich hohe Stabilitätskoeffizienten. F. M. Spinath, Ronald, Harlaar, Price und Plomin (2003) konnten z. B. im Rahmen der *Twins' Early Development Study*¹ (TEDS; Oliver & Plomin, 2007; Trouton, Spinath & Plomin, 2002) in einer Stichprobe von ca. 7000 Kindern für den *g-Faktor* im Alter von zwei bis drei Jahren eine beachtliche Stabilität von $r_{tt} = 0.69$ belegen, die im Alter von drei bis vier Jahren auf $r_{tt} = 0.71$ ansteigt. Stabilitätskoeffizienten in ähnlicher Höhe wurden auch aus der Münchner *Longitudinalstudie zur Genese individueller Kompetenzen* (LOGIK; Schneider & Stefanek, 2004; Schneider & Bullock, 2008) berichtet. Merkmale der verbalen sowie nonverbalen Intelligenz wurden im Verlauf der Studie zu insgesamt sieben Zeitpunkten erfasst. Während die Langzeitstabilität der verbalen Intelligenzwerte zwischen dem 4. und 23. Lebensjahr $r_{tt} = 0.26$ betrug, fiel die Stabilität zwischen dem 7. und 23. Lebensjahr mit $r_{tt} = 0.54$ schon deutlich höher aus. Die Zehnjahresstabilität zwischen dem 13. und 23. Lebensjahr betrug $r_{tt} = 0.66$. Für die Stabilitätsentwicklung der nonverbalen Intelligenz werden vergleichbare Werte berichtet (Schneider &

¹ Bei der britischen *Twins' Early Development Study* handelt es sich um ein groß angelegtes Längsschnittprojekt, welches die Entwicklung von Zwillingkindern ab einem Alter von zwei Jahren verfolgt.

Bullock, 2008). Dementsprechend kann ab dem mittleren Kindesalter bereits von einer hohen interindividuellen Stabilität kognitiver Fähigkeiten ausgegangen werden (z. B. Magnusson & Backtemann, 1978; Moffitt et al., 1993; F. M. Spinath et al., 2003; Sonntag, Baker & Nelson, 1958), wobei die Stabilitätskoeffizienten umso höher ausfallen, je später der erste Erhebungszeitpunkt stattfindet und je kürzer das Intervall zwischen den Messungen gewählt wird (Amelang et al., 2006; F. M. Spinath et al., 2003).

Auch in Bezug auf die intraindividuelle Stabilität zeigt sich ab dem mittleren Kindesalter eine hohe Stabilität kognitiver Fähigkeiten. Während diese bis in der frühen Kindheit noch stärkeren Schwankungen unterliegt (Holling et al., 2004; Wilson, 1983), zeigten Moffitt et al. (1993), dass innerhalb eines Zeitraumes von sechs Jahren nur ein geringer Anteil der Kinder eine nicht auf den Messfehler zurückführbare Veränderung in ihren Intelligenztestwerten aufwiesen. Diese Stabilisierung zeigte sich bereits ab einem Alter von sieben Jahren.

2.5 Abschließende Bewertung

Den bisherigen Ausführungen folgend ist ein gewisses Maß an Intelligenz zwingend notwendig, um einen bestimmten Leistungsstandard zu erreichen. Dennoch ist die Beziehung zwischen der potenziellen Leistungsfähigkeit eines Schülers und seiner letztlichen Performanz keineswegs perfekt (H. Heckhausen, 1980). Der Großteil an Varianz bleibt auch nach Berücksichtigung allgemeiner kognitiver Fähigkeiten noch unerklärt, wodurch weitere, nicht kognitive Einflussfaktoren (z. B. motivationale Voraussetzungen, Aufmerksamkeit, Ausdauer oder Selbstkontrollfähigkeiten) vor allem im pädagogischen Kontext große Beachtung finden (Helmke, 1997). Das Interesse an alternativen Einflussfaktoren richtet sich dabei insbesondere auf solche Faktoren, die im Allgemeinen als relativ leicht beeinflussbar gelten und denen aus diesem Grund – trotz zum Teil nur geringer Vorhersagebeiträge – erhebliche praktische Relevanz zugesprochen wird (Helmke & Weinert, 1997; B. Spinath et al., 2006; F. M. Spinath, Toussaint, Spengler & Spinath, 2008b). Als solche *weichen* Faktoren gelten beispielsweise motivationale Faktoren und Eigenschaften wie Stressresistenz oder soziale Kompetenz (F. M. Spinath et al., 2008b). Auch im Kontext der vorliegenden Arbeit steht die prädiktive Bedeutsamkeit motivationaler Variablen im Mittelpunkt der Betrachtung, wobei deren Vorhersagebeiträge auch in Relation zu allgemeinen kognitiven Fähigkeiten gesetzt werden.

Durch das längsschnittliche Design der KoSMoS Studie rückt zudem die Frage nach Stabilität und Veränderung schulischen Erfolgs und seiner Prädiktoren in den Fokus der durchgeführten Analysen. Für Intelligenz kann hier sowohl von einer hohen strukturellen als auch interindividuellen Stabilität ausgegangen werden, die sich einerseits

bereits in der Kindheit zeigt und sich andererseits über ein langes Zeitintervall hinweg nachweisen lässt.

Im nachfolgenden Kapitel erfolgt eine genauere Betrachtung des Prädiktors Motivation. Zu Beginn werden die relevanten theoretische Grundlagen erörtert und die hier verwendeten Motivationsmaße *Fähigkeitsselbstkonzept* und *Intrinsische Motivation* eingeführt. Hiernach erfolgt die Bewertung der prognostischen Relevanz beider Faktoren, auch unter gleichzeitiger Berücksichtigung allgemeiner kognitiver Fähigkeiten. Einen besonderen Stellenwert nehmen dabei Arbeiten ein, die sich im Rahmen von längsschnittlichen Untersuchungen mit dem Fähigkeitsselbstkonzept als Prädiktor schulischen Erfolgs befasst haben (vgl. Abschnitt 3.4.1). Abschließend wird auf Forschungsbefunde zur Stabilität der motivationalen Variablen Bezug genommen.

3 Bedingungsfaktoren schulischen Erfolgs: Motivation

Bei der Suche nach Einflussfaktoren, die neben der allgemeinen Intelligenz Varianz im schulischen Erfolg aufklären können, rücken motivationale Faktoren, die allgemein als Schlüsselfaktoren bei der Initiierung und Aufrechterhaltung von Verhalten und damit auch von schulischem Leistungsverhalten gelten, in den Mittelpunkt des Interesses. Zahlreiche Autoren betrachten motivationale Aspekte, nicht zuletzt aufgrund der ihnen zugesprochenen leichten Beeinflussbarkeit, als wichtige und einflussreiche Determinanten schulischen Erfolgs und interindividueller Unterschiede in der Leistungsentwicklung (z. B. Helmke, 1997; Helmke & Schrader, 2006; Howe, 1987; Lens, 1987). Die Zusammenhänge zwischen Motivation und Schulleistung wurden dabei vor allem im Hinblick auf die Konstrukte Leistungsmotivation, Zielorientierung, intrinsische vs. extrinsische Motivation sowie Interesse untersucht (Schiefele, 2009). Die Ergebnisse aus diesen Studien variieren zwar in Abhängigkeit der Indikatoren von Leistung (z. B. Schulnoten oder standardisierte Leistungstests), der Operationalisierung motivationaler Faktoren, der Leistungsfähigkeit innerhalb der untersuchten Stichprobe und der untersuchten Klassenstufe (Hansford & Hattie, 1982; Helmke & Weinert, 1997), bestätigen jedoch mehrheitlich die angenommene Relevanz der Motivation auch über Intelligenz hinaus (z. B. B. Spinath et al., 2006; Steinmayr & B. Spinath, 2009). Dementgegen findet sich jedoch auch die Auffassung, dass es neben der Intelligenz kein anderes psychologisches Konstrukt gibt, welches zusätzlich systematisch zur Vorhersage von Schulerfolg beitragen kann (Gagné & St Père, 2001). In diesem Zusammenhang spielt sicherlich auch eine Rolle, ob lediglich der inkrementelle Vorhersagebeitrag betrachtet (z. B. Gottfried, 1990; Schicke & Fagan, 1994; E. M. Skaalvik & S. Skaalvik, 2004), oder auf Analyseverfahren zurückgegriffen wird, mit denen auch ein potenzieller gemeinsamer Varianzanteil der Prädiktoren berücksichtigt werden kann (vgl. hierzu Abschnitt 3.4.3).

Entsprechend der Ausrichtung vorliegender Arbeit auf die Prädiktion von Schulerfolg durch kognitive und motivationale Merkmale liegt der Fokus der nachfolgenden Darstellung auf Modellen der Lern- und Leistungsmotivation. Einen umfassenden Überblick über die Motivationsforschung geben z. B. Eccles, Wigfield und Schiefele (1998), H. Heckhausen (2006) sowie Rheinberg (2006). In einem ersten Teil erfolgt eine allgemeine Begriffsbestimmung. Hieran schließt sich die Einführung des Konstruktes *Leistungsmotivation* sowie des *Erwartungs-Wert-Modells* von Jacquelynne S. Eccles und Kollegen

(Eccles et al., 1983; Wigfield & Eccles, 2000) als für diese Arbeit grundlegendes theoretisches Modell der Lern- und Leistungsmotivation an.

Im hierauf folgenden empirischen Teil wird zunächst der Forschungsstand zur prädiktiven Validität der Variablen *Fähigkeitsselbstkonzept* (FSK) sowie *Intrinsische Motivation* (IM) dargelegt, wobei Studien zur längsschnittlichen Beziehung von motivationalen Variablen und schulischem Erfolg ein besonderes Gewicht zukommt. Zuletzt erfolgt eine zusammenfassende Darstellung des Kenntnisstandes zur individuellen Entwicklungsstabilität motivationaler Variablen.

3.1 Begriffsbestimmung

In den letzten Jahrzehnten fand der Begriff *Motivation*, losgelöst von einem wissenschaftlichen Kontext, auch Eingang in unsere Alltagssprache. Im Rahmen dieses Alltagsverständnisses ist der Begriff der *Motivation*, oder spezifischer *des motivierten Verhaltens*, einerseits verbunden mit der Erklärung sozial erwünschter Verhaltensweisen, wie z. B. der freiwilligen Übernahme zusätzlicher Aufgaben. Andererseits wird er auch zur Erklärung für ein Versagen in Leistungssituationen, wie bei sportlichen Wettkämpfen oder im schulischen Kontext, herangezogen (Holodynski, 2007; Vollmeyer, 2005).

Die wissenschaftliche Auseinandersetzung in Bezug auf die Erklärung der Aufnahme und Aufrechterhaltung bestimmter Verhaltensweisen hat innerhalb der Psychologie eine lange Tradition. Umfassende Motivationstheorien und damit verbundene Terminologien liegen beispielsweise aus der Sozial-, der Pädagogischen-, der Persönlichkeits- und der Tiefenpsychologie vor (H. Heckhausen, 2006). Vor diesem Hintergrund erscheint es wenig sinnvoll, allgemein von *der* Motivation zu sprechen, denn diese lässt sich in qualitativ unterschiedliche Komponenten, die entsprechend den Modellvorstellungen der jeweiligen Forschungstradition aus einer jeweils anderen Perspektive beschrieben und erklärt werden, unterteilen (Rheinberg, 2006; Ryan & Deci, 2000a; Wild, Hofer & Pekrun, 2006). Für die vorliegende Studie sind hier insbesondere die Modellannahmen und Konzepte aus dem Kontext der Pädagogischen Psychologie von Bedeutung. Innerhalb dieser werden motivationale Merkmale und Prozesse vor allem auf das Lernen bezogen. In diesem Zusammenhang bilden die Motivation und das Interesse der Schüler eine Grundvoraussetzung für deren längerfristige Auseinandersetzung mit den Lerninhalten. Der besondere Stellenwert der Motivation ergibt sich hierbei jedoch nicht nur aus dieser leistungsförderlichen Wirkung. Zusätzlich profitieren auch bestimmte, auf den Lernprozess positiv einwirkende Verhaltensweisen, wie z. B. Tiefenverarbeitungsstrategien, von einem gewissen Ausmaß an Motivation, sich mit dem entsprechenden Thema auseinanderzusetzen (Schiefele, 1999). Schließlich ergibt sich, wie bereits er-

wähnt, die Bedeutung der Motivation auch daraus, dass motivationale Merkmale, z. B. im Vergleich zu kognitiven Fähigkeiten, als relativ leicht beeinflussbare psychische Größen gelten (B. Spinath et al., 2006).

Die angesprochene Heterogenität an Modellen und damit verbundenen Konstrukten und Terminologien führte letzten Endes auch zu einer kaum mehr überschaubaren Fülle an Publikationen aus unterschiedlichsten Forschungstraditionen, wodurch eine vergleichende Interpretation von Studienergebnissen erschwert wird (Murphy & Alexander, 2000). Um für die nachfolgenden Überlegungen begriffliche Klarheit zu schaffen, wird nach Rheinberg (2006) Motivation als eine

„aktivierende Ausrichtung des momentanen Lebensvollzugs auf einen positiv bewerteten Zielzustand [hin]“

(Rheinberg, 2006, S. 15)

verstanden. Motivation erklärt demnach die Zielgerichtetheit menschlichen Handelns und gilt dementsprechend als zentrales Element zur Erklärung von Verhalten (Vollmeyer, 2005). Sie kann als „psychische Kraft“ oder „Verhaltensbereitschaft“ verstanden werden (Schiefele, 2009), die vor allem das angestrebte *Ziel* einer Handlung (z. B. Prüfungsvorbereitung vs. Freunde treffen), die *Ausdauer* (z. B. die zur Prüfungsvorbereitung aufgewendete Zeit) sowie die *Intensität* des Verhaltens (z. B. das Ausmaß der Konzentration beim Lernen) beeinflusst (vgl. z. B. Rheinberg, 2006; Schunk, Pintrich & Meece, 2008). Nach Schiefele (2009) legt die Definition von Motivation als psychische Kraft nahe, dass es sich dabei um einen aktuellen bzw. vorübergehenden Zustand handelt. Diese Annahme geht auf den feldtheoretischen Ansatz von Kurt Lewin (1935) zurück, in dessen Zentrum die aktuelle Motivation und deren Genese stehen.

Neben aktuellen motivationalen Merkmalen, kann Motivation auch als stabiles, überdauerndes Persönlichkeitsmerkmal betrachtet werden. Diese Annahme wurzelt in instinkt- und bedürfnistheoretischen Ansätzen, die u. a. das Ziel verfolgten, menschliches Verhalten möglichst vollständig auf motivationale Dispositionen zurückzuführen (z. B. McDougall, 1937; Murray, 1938; zitiert nach Schiefele, 2009). Als Formen dispositioneller Motivation werden insbesondere Zielorientierungen, Interesse, Selbstkonzepte eigener Fähigkeiten, Formen überdauernder extrinsischer bzw. intrinsischer Motivation und sogenannte *Motive* unterschieden (Schiefele, 2009). Letztere werden als überdauernde Bewertungsvorlieben, also als stabile Präferenzen für das Erleben spezifischer Zustände, aufgefasst (Rheinberg, 2006). Im Bereich der Motivationsforschung werden primär drei Motive unterschieden: das Leistungs-, das Macht- und das Anschlussmotiv (für eine Übersicht vgl. H. Heckhausen, 2006). Verhaltenswirksam werden Motive allerdings erst, wenn sie durch bestimmte Situationsmerkmale, sogenannte Anreize, aktiviert werden. Motiv und Anreiz tragen schließlich zur Entstehung aktueller Moti-

vation bei (Rheinberg, 2006; Schiefele, 2009). Das Leistungsmotiv nimmt innerhalb der Pädagogischen Psychologie eine besondere Stellung ein und spielt insbesondere bei der Erklärung schulischer Leistungen eine wichtige Rolle. Vermutlich auch durch dessen Relevanz für den Lernbereich stellt das Leistungsmotiv das am häufigsten untersuchte Motiv dar (Brunstein & H. Heckhausen, 2006; Rheinberg, 2006).

3.2 Das Konstrukt Leistungsmotivation

Die Leistungsmotivationsforschung beschäftigt sich einerseits mit der Frage, warum manche Menschen erfolgreicher sind als andere und fragt andererseits nach den motivationalen Kräften, die Menschen dazu veranlassen, unterschiedliche Leistungsniveaus anzustreben (Gerrig & Zimbardo, 2008; Holodynski, 2007). Personen mit ausgeprägtem Leistungsmotiv zeichnen sich dabei insbesondere durch ihr Streben nach dem Erleben eigener Kompetenz und den Willen aus, sich an einem selbst gewählten Gütemaßstab zu messen (H. Heckhausen, 2006; Rheinberg, 2006; Schiefele, 2009). Dieser Gütemaßstab kann sich entweder aus dem Vergleich der eigenen aktuellen Leistung mit früheren eigenen Leistungen oder aus dem Vergleich mit den Leistungen anderer Personen ergeben (Brunstein & H. Heckhausen, 2006). Der Antrieb zur Initiierung einer Handlung im Sinne des Leistungsmotivs geht jedoch immer von der Person selbst aus (Gerrig & Zimbardo, 2008). Der Begriff *Leistungsmotivation*, wie er sich in der psychologischen Forschung etabliert hat, ist allerdings nicht mit scheinbar ähnlichen Begriffen der Alltagssprache wie Fleiß, Eifer, Arbeitswille oder Strebsamkeit gleichzusetzen (Rheinberg, 2006). Nicht jedes Bestreben etwas zu erreichen ist auch gleichzeitig leistungsmotiviert, ein bestimmtes Ziel kann durchaus auch aus anderen Gründen angestrebt werden, z. B. der Erwartung eines höheren Gehalts oder gesellschaftlichen Ansehens.

Mit Blick auf die empirische Forschung zur Leistungsmotivation lässt sich zunächst feststellen, dass sich erste Ansätze zur Beschreibung des Leistungsbedürfnisses von Personen bereits bei Henry A. Murray (Murray, 1938) finden. Entscheidend geprägt wurde die Motivationsforschung in der Folge allerdings von David C. McClelland und John W. Atkinson, die als Begründer der Leistungsmotivationsforschung gelten (Atkinson, 1957; McClelland, Atkinson, Clark & Lowell, 1953).

Im Allgemeinen konzentrieren sich leistungsbezogene Motivationskonstrukte vor allem darauf, die Auswirkungen des Leistungsmotivs auf eine Reihe von leistungsbezogenen Verhaltensmerkmalen, insbesondere die Anstrengung, die Ausdauer und die Bevorzugung herausfordernder Aufgaben, zu erklären (Brunstein & H. Heckhausen, 2006; Eccles et al., 1998).

Im Zuge dieser Überlegungen wurden zahlreiche theoretische Modelle entwickelt, die

sich jedoch mehrheitlich einer von zwei Klassen zuordnen lassen (vgl. B. Spinath et al., 2006): Solchen, die sich mit dem *Wert*, den eine Person einer bestimmten Handlung bzw. den Auswirkungen dieser Handlung zuschreibt, beschäftigen und solchen, die sich mit dieser Handlung und ihren Auswirkungen verbundenen *subjektiven Erwartungen* zuwenden. Zu Wert-Konstrukten zählen hier u.a. *Motive* (z. B. McClelland et al., 1953), *Ziele* (z. B. Dweck & Legett, 1988, Elliot & McGregor, 2001), Theorien zur *intrinsischen vs. extrinsischen Motivation* (z. B. Deci & Ryan, 1985) und *der einer Aufgabe zugeschriebene Wert* (task value; z. B. Wigfield & Eccles, 1992). Im Bereich subjektiver Erwartungen finden sich beispielsweise *Fähigkeitsselbstkonzepte* (z. B. Eccles & Wigfield, 1995; Guay, Marsh & Boivin, 2003), die *Selbstwirksamkeit* (z. B. Bandura, 1997) oder Theorien zu *Kontrollüberzeugungen* (z. B. T. D. Little, Oettingen, Stetsenko & Baltes, 1995).

Neben Theorien, die einzig einen der beiden Aspekte zur Erklärung motivationaler Prozesse einbeziehen, finden sich auch Modelle, die beide Perspektiven integrieren. Diese *Erwartungs-mal-Wert* Theorien führen im Allgemeinen das Entstehen der aktuellen Motivation (z. B. die Absicht seine Hausaufgaben zu erledigen) auf die momentanen erwartungs- und wertbezogenen Kognitionen zurück (Schiefele & Köller, 2006). *Erwartungen* beziehen sich demnach auf die Wahrscheinlichkeit, eine Handlung erfolgreich ausführen zu können, bei *Werten* handelt es sich um die subjektive Bedeutsamkeit, die der Handlung und ihren Folgen beigemessen wird (Schiefele, 2009). Insgesamt lässt sich konstatieren, dass Erwartungs-mal-Wert Ansätze in den psychologischen Theorien zur Erklärung motivierten Verhaltens eine bedeutende Stellung einnehmen. Bereits in den theoretischen Überlegungen von Lewin (1935) finden sich Annahmen späterer Erwartungs-mal-Wert Theorien. Einen bedeutenden Beitrag leistete wiederum die Arbeit von Atkinson, der ihm Rahmen der Leistungsmotivationsforschung in den 1950er- und 1960er-Jahren das Zusammenwirken von Erfolgswahrscheinlichkeit und Nützlichkeit untersuchte (vgl. z. B. Atkinson, 1957). Entscheidende modelltheoretische Modifikationen wurden in der Folge unter anderem von Jacquelynne S. Eccles und Allan Wigfield vorgenommen (siehe Abschnitt 3.3). In Erweiterung der klassischen Modelle wird die Wertkomponente allerdings weiter ausdifferenziert und durch Berücksichtigung von Komponenten wie dem *Selbstkonzept* und von *Opportunitätskosten* modifiziert (Maaz, Hausen, McElvany & Baumert, 2006; Schiefele, 2009).

Innerhalb der Pädagogischen Psychologie hat sich das bereits angesprochene *Erwartungs-Wert-Modell* der Leistungsmotivation von Eccles und Wigfield (Eccles et al., 1983; Eccles & Wigfield, 2002; Wigfield & Eccles, 2000), nicht zuletzt durch seine Komplexität sowie einer ständigen Weiterentwicklung und empirischen Bewährung, weitgehend durchgesetzt. Aufgrund dessen dient dieses Modell auch in vorliegender Arbeit als theoretischer Bezugsrahmen und wird nachfolgend zunächst zusammenfassend

beschrieben. Anschließend werden die beiden Modellkomponenten *Fähigkeitsselbstkonzept* sowie *Intrinsische Motivation*, die hier als motivationale Variablen zur Vorhersage von Schulerfolg herangezogen werden, einer differenzierten Betrachtung unterzogen.

3.3 Das Erwartungs-Wert-Modell der Leistungsmotivation sensu Eccles

Das Erwartungs-Wert-Modell der Leistungsmotivation (vgl. Abbildung 3.1), wie es von Eccles und Wigfield zu Beginn der 1980er Jahre entwickelt wurde, basiert, wie im vorherigen Abschnitt bereits eingeführt, auf den Annahmen von Atkinson. Sowohl die Erwartungs- als auch die Wertkomponente sind jedoch in diesem Ansatz weitaus differenzierter konzipiert und stehen in Bezug zu soziokulturellen sowie psychologischen Einflussfaktoren.

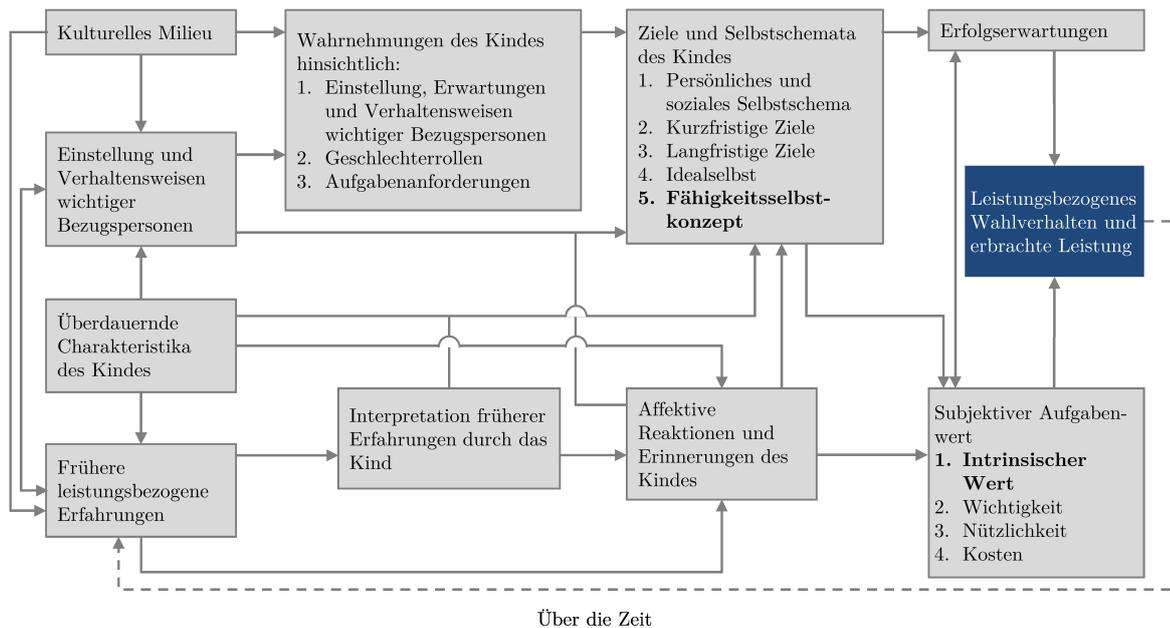


Abbildung 3.1. Das Erwartungs-Wert-Modell der Leistungsmotivation. Eigene Darstellung in Anlehnung an Abbildung 1 aus Eccles und Wigfield, 2002, S. 119.

Die *soziokulturellen Faktoren* bilden den objektiv beobachtbaren Rahmen, in dem Personen aufwachsen und dabei Erfahrungen mit Leistungssituationen sammeln. Die *psychologischen Komponenten* beziehen sich hingegen auf die subjektive Wahrnehmung der Realität durch das Kind (Eccles & Wigfield, 2002). Dementsprechend beeinflussen nicht die objektiv beobachtbaren Merkmale einer Situation die letztlich leistungsbezo-

gene Entscheidung, sondern die Art und Weise, wie ein Kind bestimmte Leistungssituationen wahrnimmt und bewertet. Entscheidungen werden also im Modell als Ergebnis der als positiv oder negativ bewerteten Aufgabencharakteristika und der mit dieser Entscheidung einhergehenden Kosten aufgefasst (Eccles & Wigfield, 2002; Maaz et al., 2006). Direkt beeinflusst wird das leistungsbezogene Wahlverhalten und die gezeigte Performanz im schulischen und akademischen Kontext durch die beiden Prädiktoren *Erfolgserwartung* sowie *subjektiver Aufgabenwert* (Eccles et al., 1983; Eccles & Wigfield, 2002; Wigfield et al., 1997). *Subjektive Erfolgserwartungen* werden definiert als die persönliche Überzeugung eines Kindes, wie gut es eine bestimmte Aufgabe in der unmittelbaren oder nahen Zukunft erfüllen kann (Eccles et al., 1983; Eccles & Wigfield, 2002). Wigfield und Eccles (2000) nehmen subjektive Erfolgserwartungen als konzeptionell vergleichbar mit der von Bandura formulierten *Selbstwirksamkeitserwartung*¹ an (vgl. z. B. Bandura, 1997).

Die *Wertkomponente* wird den Annahmen des Modells folgend wiederum in vier Komponenten unterteilt (Eccles, 2005):

1. Den *intrinsischen Wert* der einer Aufgabe beigemessen wird. Dieser bezeichnet die Befriedigung, die eine Person durch das Ausüben einer Tätigkeit erleben kann oder auch das intrinsische Interesse an einer Aktivität.
2. Die subjektiv zugeschriebene *Wichtigkeit* der Bewältigung einer bestimmten Tätigkeit oder Aufgabe.
3. Die *Nützlichkeit*, die sich auf die persönliche Einschätzung der Relevanz einer Tätigkeit oder Aufgabe für zukünftige Pläne oder Ziele bezieht (Eccles, 2005). Im Gegensatz zum intrinsischen Wert wird mit der Nützlichkeit eine extrinsische Komponente angesprochen (Deci & Ryan, 1993; Schiefele & Köller, 2006).
4. Die mit einer Aufgabe verbundenen *Kosten* die dadurch entstehen, dass durch die Entscheidung eine bestimmte Tätigkeit durchzuführen (z. B. Hausaufgaben erledigen), die Verfügbarkeit anderer Tätigkeiten (z. B. Freunde treffen) limitiert wird. Kosten können sich einerseits auf die Einschätzung beziehen, wie viel Anstrengung investiert werden muss, um eine Aufgabe zu bewältigen. Andererseits beziehen sie sich auf einen emotionalen Aspekt, beispielsweise Leistungs- oder Versagensängste (Wigfield & Eccles, 2000).

Die Erwartungs- und Wertaspekte sind ihrerseits abhängig von spezifischen Charakte-

¹ Die Selbstwirksamkeitserwartung bezeichnet die Überzeugung einer Person, das zur Erlangung eines Handlungsergebnisses erforderliche Verhalten erfolgreich ausführen zu können (Moschner & Dickhäuser, 2006).

ristika des Kindes, wie z. B. seinen Zielen und seinem generellen Selbstkonzept. Bei diesen kindspezifischen Merkmalen wird davon ausgegangen, dass sie sich dynamisch und stetig weiterentwickeln. Sie werden geprägt durch individuell wahrgenommene Erfahrungen und Interpretationen, aber auch durch die Vielfalt an Sozialisationsinflüssen. Der Einfluss sämtlicher sekundärer Faktoren (also Erfahrungen, Persönlichkeitseigenschaften, Erziehung oder auch kulturelle Einflüsse) auf leistungsbezogenes Verhalten kann demnach nur mittelbar über die Hauptfaktoren Erfolgserwartungen und Aufgabenwert erfolgen (Wigfield & Eccles, 2000).

Die beiden hier betrachteten Modellkomponenten zur Vorhersage schulischen Erfolgs werden nachfolgend näher beschrieben.

3.3.1 Erwartungskomponente: Fähigkeitsselbstkonzept

Mit dem Begriff *Selbstkonzept* wird allgemein die mentale Repräsentation einer Person über ihre eigenen Fähigkeiten und Eigenschaften beschrieben (Moschner & Dickhäuser, 2006). Diese Wahrnehmung der eigenen Fähigkeiten bildet sich im Laufe der Entwicklung aus den sozialen Erfahrungen eines Individuums durch direkte und indirekte Rückmeldungen aus seiner Umwelt (Moschner & Dickhäuser, 2006; Shavelson, Hubner & Stanton, 1976). Im Erwartungs-Wert-Modell beziehen sich *Fähigkeitsselbstkonzepte* demgemäß ebenfalls auf die persönliche Wahrnehmung der eigenen Kompetenz, aber hier insbesondere im Hinblick auf die spezifischen, aktuellen Fähigkeiten (Wigfield & Eccles, 2000). *Fähigkeitsselbstkonzepte* und *Subjektive Erfolgserwartungen* werden im Modell in enger Relation zueinander verstanden, denn aus der Einschätzung der eigenen Leistungsfähigkeit in einem bestimmten Bereich wird die Erwartung abgeleitet, auch zukünftig in diesem Bereich erfolgreich zu sein (Eccles & Wigfield, 2002; Wigfield & Eccles, 2000). Empirisch hat sich die konzeptionelle Unterscheidung beider Aspekte allerdings nicht bestätigt, weshalb sie in zahlreichen Studien als *ein* Konstrukt erhoben werden (Eccles & Wigfield, 2002).

Das große wissenschaftliche Interesse am *Selbstkonzept* ist vor allem dadurch zu erklären, dass in zahlreichen psychologischen Disziplinen (z. B. der Gesundheits-, Entwicklungs-, Sozial- und Pädagogischen Psychologie) einem positiven Selbstkonzept eine wichtige und wünschenswerte Funktion zugesprochen wird (Marsh & Martin, 2011). Diesen Stellenwert erhält das Selbstkonzept zum einen durch seine angenommene enge Verbindung zum Verhalten, sowie zu emotionalen und kognitiven Vorgängen (z. B. Ängstlichkeit, akademische Leistungen oder Zufriedenheit; Branden, 1994). Zum anderen wird angenommen, dass sich das Selbstkonzept in einer frühen Entwicklungsphase ausbildet, dann aber relativ stabil bleibt und überdauernd ist (Eder & Mangelsdorf,

1997). Aus diesem Grund nimmt die Förderung eines positiven Selbstkonzeptes auch eine zentrale Rolle in den Zieldefinitionen von Frühförderprogrammen sowie Erziehung und Unterricht ein (vgl. Marsh & Craven, 2006). Im Rahmen der Pädagogischen Psychologie wurden Selbstkonzepte vor allem im Hinblick auf deren Bedeutung für den Schulerfolg intensiv beforscht (zu den Ergebnissen dieser Studien vgl. Abschnitt 3.4.1).

Eine häufig gestellte Frage innerhalb dieser Forschungstradition betrifft die anzunehmende *Globalität* vs. *Spezifität* des Selbstkonzeptes. Während in früheren Arbeiten ein unidimensionales Modell ähnlich Spearman's Generalfaktorentheorie der Intelligenz angenommen und ein globales Selbstwertgefühl erfasst wurde, werden heute eher multidimensionale Ansätze favorisiert (Marsh & Craven, 2006; Marsh & Martin, 2011; Möller & Trautwein, 2009). Das bekannteste und am weitesten verbreitete Modell stellt hier das von Shavelson et al. (1976) vorgeschlagene multifaktorielle, hierarchische Modell des Selbstkonzeptes dar. Auf der generellsten Ebene steht das allgemeine Selbstkonzept, auf der darunterliegenden Ebene werden ein akademisches sowie nicht-akademische Selbstkonzepte (das Soziale, das Emotionale und das Psychische) unterschieden. Das hier interessierende *akademische Selbstkonzept* bezieht sich auf die generalisierten fachspezifischen Fähigkeitseinschätzungen und lässt sich auf einer weiteren Ebene in verschiedene schulfachspezifische Selbstkonzepte untergliedern (Brunner et al., 2010; Möller & Köller, 2004; Schöne, Dickhäuser, Spinath & Stiensmeier-Pelster, 2003). Diese sind im Rahmen pädagogisch-psychologischer Fragestellungen von besonderer Relevanz und werden allgemein als *Fähigkeitsselbstkonzepte* bezeichnet (Moschner & Dickhäuser, 2006). Aktuelle Forschungsarbeiten lassen zudem auf eine weitere Unterteilung bereichsspezifischer Fähigkeitsselbstkonzepte in eine *kognitive* und eine *affektive* Komponente schließen (Arens, Yeung, Craven & Hasselhorn, 2011).

Innerhalb des Erwartungs-Wert-Modells von Eccles und Kollegen (Eccles et al., 1983; Eccles & Wigfield, 2002; Wigfield & Eccles, 2000) wird ebenfalls von einer hohen Spezifität des Fähigkeitsselbstkonzeptes ausgegangen und dieses entsprechend als schulfachspezifisches Fähigkeitsselbstkonzept operationalisiert. Marsh, Debus und Bornholt (2005a) konnten beispielsweise bereits für Kinder unter acht Jahren eine multidimensionale Struktur des Fähigkeitsselbstkonzeptes nachweisen. Für eine bereichsspezifische und damit multidimensionale Erfassung von Fähigkeitsselbstkonzepten spricht auch eine niedrige und zum Teil leicht negative Korrelation zwischen mathematischem und verbalem Selbstkonzept einerseits und eine Korrelation nahe Null zwischen globalem Selbstkonzept und schulischen Leistungsmessungen andererseits (Schulnoten, Schulleistungstests; Marsh, Trautwein, Lüdtke, Köller & Baumert, 2006).

3.3.2 Wertkomponente: Intrinsische Motivation

In der Literatur wird *intrinsische Motivation* üblicherweise als der Wunsch oder die Absicht definiert, eine bestimmte Handlung durchzuführen, weil die Handlung selbst als interessant, spannend, herausfordernd usw. wahrgenommen wird (Schiefele & Köller, 2006). Die Gründe für die Durchführung einer Handlung sind dementsprechend im Handlungsvollzug selbst zu suchen. Im Gegensatz dazu liegt *extrinsisch* motiviertem Verhalten der Wunsch oder die Absicht zu Grunde, positive Folgen herbeizuführen oder negative Folgen zu vermeiden (Schiefele & Köller, 2006). Ferner treten diese Verhaltensweisen in der Regel nicht spontan auf (Deci & Ryan, 1993).

Wigfield und Eccles (2000) sehen ihre Definition intrinsisch motivierten Verhaltens (vgl. Abschnitt 3.3) in enger Relation zur Konzeption von Deci und Ryan, die in ihrer *Selbstbestimmungstheorie* verschiedene Ansätze zur intrinsischen Motivation integrieren (Deci & Ryan, 1985; Ryan & Deci, 2000b; Ryan & Deci, 2000a). Intrinsisch motiviertes Verhalten ist dieser Theorie zufolge nur dann möglich, wenn sich die Person als kompetent und selbstbestimmt erlebt (Deci & Ryan, 1985; Ryan & Deci, 2000b; Ryan & Deci, 2000a; Schiefele & Köller, 2006). Allerdings können bestimmte Formen extrinsischer Motivation ebenfalls als selbstbestimmt erlebt werden (Deci & Ryan, 1993).

Bezogen auf den schulischen Kontext drückt sich intrinsische Motivation als Freude am Lernen aus und umfasst Aspekte wie Neugier, Durchhaltevermögen und den Willen, neue und eventuell schwierige Aufgaben zu erlernen (Gottfried, 1990).

Eine konzeptionelle Nähe ist weiterhin zu dem Konstrukt des *Interesses* zu sehen, das vor allem in der jüngeren Vergangenheit im Hinblick auf den schulischen Erfolg untersucht wurde (vgl. Abschnitt 3.4.2). Beiden Konzepten ist gemein, dass der Grund eines aktiven Handelns in der Aufgabe an sich zu finden ist und keine externen Konsequenzen antizipiert werden oder vorhanden sein müssen (B. Spinath & F. M. Spinath, 2005b).

3.3.3 Abschließende Bewertung

Zusammenfassend betrachtet liefert die Integration von Erwartungs- und Wertkonstrukten einen fruchtbaren Ansatz zur Untersuchung des Vorhersagebeitrags motivationaler Faktoren auf den Schulerfolg von Grund- und Sekundarschülern. Das dieser Arbeit zugrunde liegende Erwartungs-Wert-Modell der Leistungsmotivation von Eccles und Kollegen wurde in zahlreichen Studien, sowohl für Schüler der Elementar- als auch der Sekundarstufe, empirisch überzeugend belegt und bietet daher sowohl theore-

tisch als auch empirisch eine fundierte Grundlage (vgl. z. B. Eccles et al., 1983, 1993a; Eccles & Wigfield, 2002; Eccles, 2005; Wigfield & Eccles, 2002).

Beispielsweise können bereits Kinder der ersten Klasse zwischen den Fähigkeitsselbstkonzepten und dem einer Aufgabe beigemessenen Wert differenzieren (Eccles et al., 1993b). Im Unterschied zu den im Modell angenommenen Beziehungen zeigte sich empirisch, dass der subjektive Aufgabenwert am stärksten mit der *Wahl* ein bestimmtes Fach beizubehalten zusammenhängt und weniger stark mit dem leistungsbezogenen Verhalten per se (Eccles & Wigfield, 2002; Wigfield & Eccles, 2000). Weiterhin hat sich gezeigt, dass Kinder im Grundschulalter noch nicht zwischen den Wertaspekten *Nützlichkeit* und *Wichtigkeit* differenzieren können (Wigfield & Eccles, 2000), weshalb im Kontext der vorliegenden Arbeit einzig der Wertaspekt *Intrinsische Motivation* betrachtet wird.

Nachfolgend werden das *Fähigkeitsselbstkonzept* und die *intrinsische Motivation* hinsichtlich ihrer prognostischen Relevanz zur Vorhersage schulischen Erfolgs betrachtet. Hierauf folgt eine zusammenfassende Darstellung der empirischen Befundlage zur individuellen Entwicklungsstabilität der Variablen.

3.4 Vorhersage von Schulerfolg durch motivationale Variablen

Wie oben bereits angeführt wurde die prognostische Relevanz verschiedener motivationaler Variablen für schulischen Erfolg empirisch vielfach bestätigt (z. B. Bouffard, Marcoux, Vezeau & Bordeleau, 2003; Helmke, 1993; Schiefele & Schreyer, 1994; Valentine, DuBois & Cooper, 2004). In frühen Studien von Bloom (1976) sowie Cattell, Barton und Dielman (1972) werden zwischen 20% und 25% Varianzaufklärung durch verschiedene Motivationsvariablen berichtet, Metaanalysen zufolge sind diese Werte allerdings nach unten zu korrigieren: Hansford und Hattie (1982) geben hier eine mittlere Korrelation von 0.21 und einen aufgeklärten Varianzanteil von 4% bis 7% an. In Studien, die neben kognitiven auch motivationale Prädiktoren berücksichtigen, wurde zudem ein inkrementeller Vorhersagebeitrag verschiedener Motivationskonstrukte bestätigt (z. B. Gottfried, 1990; Lloyd & Barenblatt, 1984; Sauer & Gattringer, 1986; B. Spinath et al., 2006).

Anknüpfend an die vorangegangene theoretische Einführung werden im Folgenden zunächst Ergebnisse aus Studien vorgestellt, die sich mit dem Vorhersagebeitrag des FSKs beschäftigen. Ein besonderes Gewicht nehmen hierbei die Arbeiten der Arbeitsgrup-

pe um Herbert W. Marsh ein, die in zahlreichen Studien den Zusammenhang zwischen akademischem Selbstkonzept und Schulerfolg untersucht und auf Basis dieser Ergebnisse das so genannte *Reciprocal Effects Model* (REM) abgeleitet haben. Dieses Modell dient in der vorliegenden Arbeit als Grundlage der Analyse des Vorhersagebeitrags der motivationalen Variablen FSK und IM unter Berücksichtigung der allgemeinen kognitiven Fähigkeit und wird hier im Rahmen eines längsschnittlichen Multikohortendesigns umgesetzt.

Hieran schließt sich eine zusammenfassende Betrachtung von Studienergebnissen zu IM sowie dem inhaltlich verwandten Konzept des Interesses an. Abschließend werden Studien berichtet, in denen mehrere Prädiktoren simultan berücksichtigt wurden.

3.4.1 Prognostische Relevanz des Fähigkeitsselbstkonzeptes

Fähigkeitsselbstkonzepte zählen zweifelsfrei zu den am häufigsten untersuchten motivationalen Merkmalen im Hinblick auf deren prädiktive Validität für schulischen Erfolg. Ausgehend von den Ergebnissen zahlreicher empirischer Studien kann ein mittlerer bis hoher positiver Zusammenhang zwischen FSK und schulischer Leistung für Kinder unterschiedlicher Altersbereiche als bestätigt angesehen werden (z. B. B. M. Byrne, 1996; Guay et al., 2003; Hattie, 1992; E. M. Skaalvik & Valas, 1999; Steinmayr & B. Spinath, 2009). Leistungsbezogenes Selbstkonzept sagt zudem auch unter Berücksichtigung der früheren Schulleistung, des familiären sozioökonomischen Status (SES) und der Familienstruktur den erreichten Bildungsgrad 10 Jahre später vorher (Guay, Larose & Boivin, 2004).

Ergebnisse aus jüngeren Überblicksarbeiten belegen konsistent einen deutlich positiven Zusammenhang zwischen schulischem Selbstkonzept und der schulischen Leistung, operationalisiert durch Fachnoten bzw. Schulleistungstests (Marsh & Craven, 2006; Robbins et al., 2004; Valentine et al., 2004). In einer Metaanalyse mit über 50 längsschnittlichen Studien zeigen Valentine et al. (2004) beispielsweise, dass sich, auch nach Kontrolle früherer Leistung, mit einem gemittelten standardisierten Pfadgewicht von 0.08 ein günstiger, wenn auch relativ geringer Einfluss durch positive Selbstüberzeugungen findet. Die Befunde weisen ferner darauf hin, dass die Effekte höher ausfallen, wenn die Erfassung der Selbstüberzeugungen bereichsspezifisch erfolgt (siehe hierzu auch Helmke & Weinert, 1997) und sie mit Noten aus dem entsprechenden Schulfach kombiniert werden. Dieser höhere Effekt bei schulfachspezifischer Erfassung in Kombination mit Schulnoten als Leistungskriterium wurde auch in verschiedenen Querschnittsanalysen nachgewiesen (z. B. Schicke & Fagan, 1994; Steinmayr & B. Spinath, 2009). Als Ursache dieser Beeinflussung der Höhe der gefundenen Zusammenhänge durch das verwendete Leistungsmaß (Schulnoten und Lehrerratings vs. standardisierte Testscores)

wird von Hansford und Hattie (1982) ein gleicher Bezugsrahmen für die Leistungseinschätzung der Lehrer sowie der Wahrnehmung eigener Fähigkeit auf Seiten der Schüler angeführt. Sowohl Lehrerurteile als auch Noten beziehen sich auf die relative Position des Schülers innerhalb des Klassenverbandes. Die Schüler nehmen ihre Einschätzung der eigenen Fähigkeit ebenfalls im Vergleich zu ihren Peers vor, zusätzlich erhalten sie im Schulalltag Rückmeldungen bzgl. ihrer Leistung durch den Lehrer.

Neben der Frage eines generellen Zusammenhangs zwischen Selbstkonzept und Leistung nahm die Frage nach der Verursachungsrichtung zwischen beiden Variablen, also ob das FSK die Ursache späterer Leistung oder vielmehr das Ergebnis früherer Leistungsbeurteilungen (oder beides) darstellt, eine Schlüsselposition ein. Gäbe es einen solchen kausalen Effekt des FSKs, würde dies implizieren, dass sich die Förderung des Selbstkonzeptes auch auf solche Variablen günstig auswirken würde, die mit dem Selbstkonzept in Beziehung stehen. Hier wurden von Calsyn und Kenny (1977) zwei Ansätze gegenübergestellt: Der *self-enhancement* Ansatz nimmt an, dass das FSK die Leistung einer Person beeinflusst und leitet davon die Annahme ab, dass eine Steigerung des Selbstkonzeptes auch eine Veränderung nachfolgender schulischer Leistung bewirkt (B. M. Byrne, 1996; Helmke & van Aken, 1995; Schöne et al., 2003). Eine Bestätigung dieser Beeinflussungsrichtung würde dementsprechend eine Berechtigung für Interventionsprogramme liefern, die auf die Förderung eines positiven Selbstkonzeptes fokussieren. Im Gegensatz dazu geht der *skill-development* Ansatz davon aus, dass fachbezogene Selbstkonzepte von schulischen und außerschulischen Leistungsrückmeldungen bestimmt werden, dass Leistungen also als ursächlich für Selbstkonzepte angenommen werden können (vgl. Möller & Trautwein, 2009; Schöne et al., 2003).

Aus theoretischer Sicht stellt sich hier allerdings weniger die Frage, ob vorangegangene Leistungsrückmeldungen bei der Etablierung akademischer Selbstkonzepte überhaupt eine Rolle spielen, denn diese Verbindung wird in vielen Selbstkonzepttheorien explizit als maßgeblich angenommen (vgl. Abschnitt 3.3.1; Marsh & Martin, 2011). Andererseits hat sich auch gezeigt, dass die Zensuren in mathematischen und sprachlichen Fächern zwar positiv miteinander korrelieren, die fachspezifischen Selbstkonzepte allerdings keine bedeutsame Korrelation miteinander aufweisen (z. B. Marsh, 1986). Dies lässt den Schluss zu, dass die Entwicklung von Fähigkeitsselbstkonzepten nicht allein auf eine Internalisierung von Leistungsrückmeldungen zurückgeht (Möller & Trautwein, 2009; B. Spinath et al., 2006). Selbstkonzepte basieren zwar teilweise auf konkreten Leistungsrückmeldungen, in der Folge werden allerdings auch soziale und interne Vergleiche zwischen den Leistungen in unterschiedlichen Schulfächern angestellt, was sich ebenfalls auf die Höhe der Fähigkeitsselbstwahrnehmung auswirkt (Möller & Trautwein, 2009; B. Spinath et al., 2006). Diese Prozesse werden zusammenfassend im sog. *Referenzrahmen-Modell* beschrieben (vgl. hierzu Abschnitt 3.5). Die eigentlich kritische Frage ist dementsprechend, ob sich eine Beeinflussung der schulischen Leistung

durch vorangegangenes Selbstkonzept im Sinne des *self-enhancement* Ansatzes auch empirisch nachweisen lässt (Marsh & Martin, 2011).

Als richtungsweisend bei der Klärung der Frage nach der kausalen Prädominanz von Selbstkonzept und Schulleistung können, nicht zuletzt durch deren präzise und methodisch differenzierte Umsetzung, die Arbeiten von Herbert W. Marsh und Kollegen angesehen werden. Aus diesen kann der Schluss gezogen werden, dass eine vereinfachende Annahme einer Kausalbeziehung im Sinne des *self-enhancement* oder des *skill-development* Ansatzes zu kurz greift. Vielmehr wächst die empirische Bestätigung für eine wechselseitige (reziproke) Beeinflussung von Selbstkonzept und Leistung (vgl. z. B. Guay et al., 2003; Marsh & Yeung, 1997; Marsh, Byrne & Yeung, 1999, 2005b).

Da die methodische Umsetzung dieses reziproken Modells und die daraus abgeleiteten Schlussfolgerungen für die vorliegende Arbeit maßgeblich sind, werden diese nachfolgend einer detaillierten Betrachtung unterzogen.

Das Reciprocal Effects Model

Zur empirischen Bewertung der kausalen Verursachungsrichtung zwischen akademischen Selbstkonzepten und schulischer Leistung wurde das in Abbildung 3.1 dargestellte prototypische REM entwickelt, mit dem sowohl das *self-enhancement*, das *skill-development* und das Modell einer *reziproken Beeinflussung* überprüft werden können (Marsh et al., 1999; Marsh & Craven, 2006). Die zu prüfenden methodischen Voraussetzungen zur Durchführung einer REM Studie wurden von Marsh et al. (1999) explizit formuliert und werden in Kapitel 9.3 dargestellt. Das generelle Vorgehen zur Prüfung der drei Modelle wird nachfolgend erläutert.

Wie aus Abbildung 3.1 hervorgeht, sollten als Grundvoraussetzung mindestens drei aufeinanderfolgende Erhebungszeitpunkte realisiert werden (Z_1 , Z_2 und Z_3). Innerhalb jeden Zeitpunktes wird eine Korrelation zwischen den latenten Variablen angenommen. Diese sind für Z_1 als reine Kovarianz zu interpretieren, in Z_2 und Z_3 als residuale Kovarianz. Außerdem wird eine Korrelation zwischen den gleichen Indikatoren zu unterschiedlichen Zeitpunkten angenommen (engl. *correlated uniqueness*, CU). Die zu prüfenden Modelle nehmen übereinstimmend eine substanzielle Stabilität der Konstrukte über die Zeit hinweg an, die in Abbildung 3.1 als durchgezogene, horizontale graue Pfeile zwischen den Zeitpunkten dargestellt ist.

Zur Prüfung der Modelle muss nun die statistische Signifikanz der Kreuzpfade (vorangegangener Schulerfolg (SE) auf nachfolgendes FSK sowie vorangegangenes FSK auf nachfolgenden SE) geprüft werden. Das *self-enhancement* Modell wird bestätigt, wenn signifikant positive Pfade lediglich in der Beziehung zwischen früherem FSK und

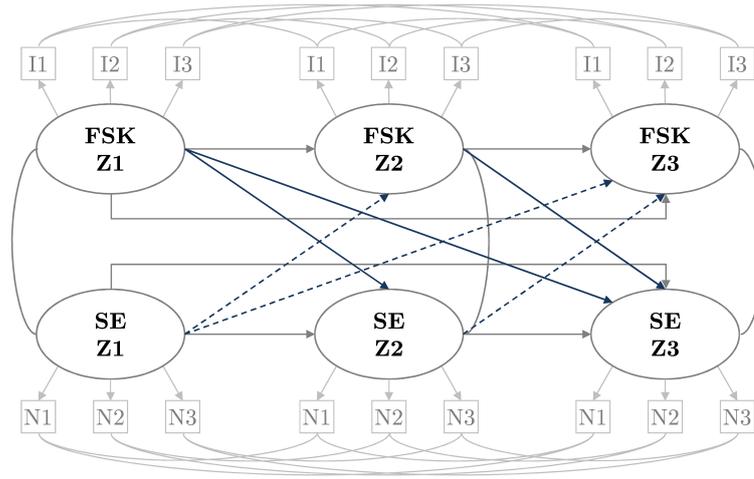


Abbildung 3.1. Prototypisches Modell zur Überprüfung der kausalen Wirkrichtung zwischen Fähigkeitsselbstkonzept und schulischem Erfolg. Eigene Darstellung in Anlehnung an Abbildung 1 aus Marsh et al., 1999; Marsh und Craven, 2006, S. 162. In diesem Modell (auch als „Full Forward Modell“ bezeichnet) werden mehrere Indikatoren für FSK und den SE zu mehreren aufeinanderfolgenden Zeitpunkten erhoben (Z1, Z2 und Z3). Die Indikatoren sind durch Vierecke dargestellt, Kreise repräsentieren die latente Schätzung von FSK und SE. Gerichtete Pfeile stehen für eine kausale Beziehung, gebogene Pfeile für eine Kovarianz.

nachfolgendem SE vorliegen (durchgezogene blaue Pfeile in Abbildung 3.1). Das *skill-development* Modell wäre bestätigt, wenn nur die Pfade zwischen vorangegangenen SE und nachfolgendem FSK signifikant positiv ausfallen (gestrichelte blaue Pfeile in Abbildung 3.1). Finden sich in beide Richtungen bedeutsame positive Effekte, gilt das REM als bestätigt. Signifikant positiv sollen dabei in allen Modellen in erster Linie die Beziehungen zwischen den zwei aufeinanderfolgenden Messzeitpunkten (Z1 auf Z2 sowie Z2 auf Z3) ausfallen. Für die Beziehung der Konstrukte zwischen Z1 und Z3 wird angenommen, dass diese zu einem großen Teil über Z2 vermittelt wird, so dass der totale (direkter plus indirekter) Effekt insgesamt positiv ausfällt (Marsh & Craven, 2006).

Marsh (1990) schließt aus einer frühen Studie (insgesamt vier Erhebungszeitpunkte im Abstand von je einem Jahr beginnend mit der 10. Klasse) zur Überprüfung der kausalen Wirkrichtung zwischen FSK und SE anhand des in Abbildung 3.1 dargestellten Modells, dass ein reines *skill-development* Modell nicht bestätigt werden kann. Vielmehr wurden auch Effekte des FSKs auf den SE ein Jahr bzw. zwei Jahre später gefunden. Dies wird von Marsh (1990) als Bestätigung einer wechselseitigen Beeinflussung und damit als Hinweis für die Adäquatheit des REM gewertet.

Als ausschlaggebend für die mittlerweile breite Akzeptanz des REM sind die Ergebnisse der unter 3.4.1 bereits erwähnten Metaanalyse von Valentine et al. (2004) zu bewer-

ten. Die Autoren finden in 90% der einbezogenen Studien einen hoch signifikanten, positiven Effekt von früherem Selbstkonzept auf nachfolgende Leistung, was Valentine et al. (2004) als klaren Beleg für das REM werten. Darüber hinaus wurde die Generalisierbarkeit des REM durch die Berücksichtigung vielfältiger Moderatoren (z. B. Alter der Schüler, Herkunftsland der Stichprobe, Operationalisierung der Variablen, Art der Modellierung) bestätigt. Lediglich die Globalität bzw. Spezifität der Selbstkonzeptmessung hatte einen signifikanten moderierenden Einfluss auf die Höhe der Effekte. Weiterhin fanden sich stärkere Effekte, wenn innerhalb der in den Studien betrachteten Zeitspanne kein Schulwechsel stattfand (Valentine et al., 2004).

Die Annahme einer reziproken Beeinflussung von Selbstkonzept und Leistung findet sich auch in vielen anderen Motivationstheorien (z. B. Bandura, 1997; B. M. Byrne, 1996; Eccles & Wigfield, 2002; Harter, 1999). Im Erwartungs-Wert-Modell von Eccles und Kollegen (Eccles et al., 1983; Eccles & Wigfield, 2002; Wigfield & Eccles, 2000) wird beispielsweise angenommen, dass sich akademische Selbstkonzepte durch vorangegangene akademische Erfolge bilden und wiederum nachfolgende akademische Leistungen direkt oder indirekt beeinflussen (Marsh & Martin, 2011).

In einigen Studien wurde das REM zusätzlich unter einer entwicklungsorientierten Perspektive untersucht. Hierbei stand die Frage im Mittelpunkt, ob das vorrangig an Stichproben älterer Kinder entwickelte reziproke Modell auch für jüngere Kinder Gültigkeit besitzt.

Valentine et al. (2004) leiteten diesbezüglich auf Basis theoretischer Überlegungen von Harter (1999) sowie Ergebnissen verschiedener empirischer Studien (z. B. Hattie, 1992; E. M. Skaalvik & Hagtvet, 1990) die Annahme ab, dass vor allem in niedrigen Klassenstufen stärkere *skill-development* als *self-enhancement* Effekte zu finden seien, wohingegen in höheren Klassenstufen reziproke Effekte vorherrschen sollten. Diese Annahme hat sich jedoch nicht bestätigt: Die Höhe der Effekte variiert nicht in Abhängigkeit des Alters oder der Klassenstufe (Valentine et al., 2004). Die Autoren schlussfolgern, ebenso wie Marsh et al. (1999), dass keine hinreichende Evidenz für eine altersabhängige Gültigkeit des REM vorliegt und dass es an weiteren, methodisch stärkeren Untersuchungen fehlt, um diese Frage abschließend beantworten zu können. Eine solche, methodisch aufwändige Studie wurde von Guay et al. (2003) umgesetzt, die zur Beantwortung dieser Frage ein Multi-Kohorten-Sequenz-Design (drei Kohorten – zweite, dritte und vierte Klasse – in drei aufeinanderfolgenden Jahren) umgesetzt haben. Ein solches *Multi-Kohorten-Sequenz-Design* (MKSD) kombiniert die Vorteile von querschnittlichen (Betrachtung mehrerer Altersgruppen) und längsschnittlichen (Betrachtung mehrerer Zeitpunkte) Forschungsansätzen und liefert damit einen geeigneten me-

thodischen Zugang zur Entkopplung von Kohorten²- und Entwicklungseffekten³ (vgl. Kapitel 9.3; Marsh & Craven, 2006; Marsh & Martin, 2011). Insgesamt betrachtet wurde die Selbstkonzeptmessung mit steigendem Alter zwar sowohl reliabler als auch stabiler und korrelierte zudem höher mit akademischer Leistung, die gefunden Unterschiede zwischen den Altersgruppen fielen jedoch äußerst gering aus. Unabhängig vom Alter der Kinder fand sich sowohl ein Effekt der Leistung auf das Selbstkonzept (*skill-development* Ansatz), als auch ein Einfluss des Selbstkonzeptes auf die Schulleistung (*self-enhancement* Ansatz) mit Pfadgewichten zwischen 0.05 und 0.26. Auf Grundlage dieser Ergebnisse sehen die Autoren das REM in verschiedenen Altersgruppen als bestätigt an.

Dennoch sollte nicht unberücksichtigt bleiben, dass auch in dieser Studie die Pfade der Effekte von der schulischen Leistung auf das akademische Selbstkonzept zwischen T1 und T2 höher ausfielen als in umgekehrter Richtung (0.26 vs. 0.18 in Kohorte 1 und 2; 0.26 vs. 0.22 in Kohorte 3). Dies spricht in der Tendenz eher für den *skill-development* Ansatz in jüngeren Altersgruppen, zumal sich die Pfadgewichte in Kohorte 3 am wenigsten stark unterscheiden. Zum gleichen Ergebnis kommt auch eine Studie von Helmke und van Aken (1995), die ebenfalls unter Rückgriff auf Strukturgleichungsmodelle in einer Stichprobe von Grundschulkindern⁴ das Vorherrschen der Beeinflussungsrichtung Schulleistung auf Selbstkonzept nachweisen konnten.

3.4.2 Prognostische Relevanz intrinsischer Motivation

Ebenso wie beim Merkmal FSK belegen empirischen Studien eine zwar meist niedrige aber konsistent positive Relation zwischen IM und schulischen Leistungen (z. B. Gottfried, 1990; Lloyd & Barenblatt, 1984; Ratelle, Guay, Vallerand, Larose & Sénécal, 2007); siehe auch Übersicht von Schiefele & Schreyer, 1994). Diese Zusammenhänge wurden sowohl für Kinder im Grundschulalter (Gottfried, 1990; Helmke, 1993; E. M. Skaalvik & S. Skaalvik, 2004; B. Spinath et al., 2006) als auch für Kinder im Sekundarschulalter (Marsh et al., 2005b; Lloyd & Barenblatt, 1984; Steinmayr & B. Spinath, 2009) in ähnlicher Höhe ($r = 0.17 - 0.63$) nachgewiesen. Die zum Teil beträchtlichen Schwankungen in den berichteten Korrelationen lassen sich vor allem auf eine unterschiedliche Operationalisierung von IM (global vs. bereichsspezifisch) sowie des Kri-

² Bei einem Kohorteneffekt bestehen Unterschiede zwischen den Entwicklungsverläufen unterschiedlicher Geburtskohorten (Lohaus, Vierhaus & Maass, 2010).

³ Als Entwicklung werden relativ überdauernde Veränderungen des Erlebens und Verhaltens über die Zeit hinweg definiert (Lohaus et al., 2010).

⁴ Die Daten entstammen der Studie zu *Schulisch organisierten Lernangeboten und der Sozialisation von Talenten, Interessen und Kompetenzen* (SCHOLASTIK), in der individuelle Entwicklungsverläufe während der Grundschulzeit in Abhängigkeit von affektiven und kognitiven Eingangsbedingungen, sowie dem schulischen Kontext untersucht wurden (Helmke, 1997).

teriums SE (Schulnoten vs. Leistungstests) zurückführen. Vergleichbar mit dem FSK finden sich auch hier in der Regel höhere Zusammenhänge bei einer domänenspezifischen Erfassung der IM und der Verwendung entsprechender Fachnoten als Schulleistungskriterium (Gottfried, 1990; Helmke & Schrader, 2006; E. M. Skaalvik & S. Skaalvik, 2004).

Schiefele und Schreyer (1994) fassen in einer Metaanalyse Studienergebnisse zur IM und deren Zusammenhang zu unterschiedlichen Indikatoren des Lernverhaltens (Schulnoten, Leistungstests, spezifische Lernkriterien) zusammen. Über die betrachteten Studien hinweg finden sich durchweg positive, wenn auch niedrige Zusammenhänge zwischen IM und verschiedenen Lernstrategien ($r = 0.30 - 0.44$), Noten ($r = 0.21$) und Leistungstests ($r = 0.24$), die in fast allen Fällen, in denen eine Berücksichtigung von Kontrollvariablen erfolgte, bestehen blieben. Durchschnittlich konnten ca. 4% der Leistungsvarianz durch IM aufgeklärt werden, wobei sich die intrinsische Lernmotivation vor allem auf konzeptuelle Formen des Lernens günstig auszuwirken scheint (Schiefele & Schreyer, 1994).

Den berichteten positiven Zusammenhängen zwischen IM und Schulleistung stehen Ergebnisse einer Studie von Gagné und St Père (2001) entgegen, in der keinerlei zusätzliche Varianzaufklärung durch IM über Intelligenz hinaus geleistet werden konnte. Die Autoren stellen hieraus abgeleitet die Relevanz motivationaler Faktoren für schulischen Erfolg grundsätzlich in Frage. Allerdings sollten die Ergebnisse dieser Studie nicht überbewertet werden. So argumentieren Steinmayr und B. Spinath (2009), dass diese Studie erhebliche methodische Mängel aufweist (eingeschränkte Varianz intrinsischer Motivation in einer reinen Mädchenschule, psychometrische Unzulänglichkeiten des Messinstrumentes, Operationalisierung des Kriteriums durch eine gepoolte Note aus vier Fächern) und die Ergebnisse aus zahlreichen anderen Studien deutlich für einen zwar geringen, aber dennoch bedeutsamen Einfluss der IM sprechen.

Empirisch finden sich jedoch vor allem Untersuchungen im Rahmen des theoretisch zur IM verwandten Konstruktes *Interesse* und dessen Relation zu schul- und fachbezogenen Wahlentscheidungen (z. B. Krapp, 1999, 2006; Schiefele, Krapp & Schreyer, 1993). Auch aus diesen Studien können Rückschlüsse auf den zu erwartenden Zusammenhang zwischen IM und SE gezogen werden.

In einer Übersichtsarbeit stellen Schiefele et al. (1993) fest, dass die Höhe des Interesses hinsichtlich eines Schulfachs in mittlerem Ausmaß mit den entsprechenden Leistungen bzw. Noten in diesem Schulfach korreliert – ein Befund, der auch in neueren Arbeiten Bestätigung findet (z. B. Barron & Harackiewicz, 2001; Harackiewicz, Barron, Tauer & Elliot, 2002; Marsh et al., 2005b). Auf Basis von Daten der PISA-Studie konnten Chiu und Xihua (2008) zeigen, dass das Interesse an Mathematik in 80% aller teilnehmen-

den Länder auch bei Kontrolle einer Reihe von Einflussvariablen (z. B. Leistungsniveau, Selbstkonzept) signifikant zur Vorhersage mathematischer Kompetenz beiträgt.

Köller, Baumert und Schnabel (2001) schlussfolgern aus den Ergebnissen ihrer Längsschnittanalyse (drei Erhebungszeitpunkte, 7. Klasse, 10. Klasse, 12. Klasse), dass in den unteren Schulstufen die Motivation der Schüler in erster Linie durch extrinsische Anreize und Werte reguliert wird. Als Auslöser zur Initiierung und Aufrechterhaltung von Verhalten sollen in dieser Zeit noch primär schriftliche Tests und Verstärkungen durch die Eltern fungieren und weniger das eigene Interesse. In den höheren Klassenstufen gewinnen die Schüler hingegen immer mehr Möglichkeiten zur Selbstbestimmung (extrinsische Anreize nehmen ab), folglich gewinnt das Interesse einen größeren Einfluss auf die Regulation des Lernverhaltens. Dies zeigt sich in den Daten einerseits in einem signifikanten direkten Effekts ($= 0.19$) des Interesses in der 10. Klasse auf die Leistungen zwei Jahre später und andererseits in der Häufigkeit, Mathematik als Leistungskurs zu wählen (Köller et al., 2001). Insbesondere der letztgenannte Befund steht im Einklang mit den Annahmen des Erwartungs-Wert-Modells von Eccles und Kollegen. Diese konnten zeigen, dass sich Merkmale der intrinsischen Motivation stärker auf akademische Wahlentscheidungen (z. B. Kurswahlen, Studienfachwahlen) denn auf die Höhe der erbrachten Leistung per se auswirken (Eccles et al., 1983; Eccles, 2005; Wigfield & Eccles, 2000). Die Autoren argumentieren, dass aufgrund der nur geringen Wahlmöglichkeiten im Bereich der Grundschule der Zusammenhang zwischen Erwartungsaspekten und Leistung deutlich höher ausfallen sollte als mit Wertaspekten (Eccles et al., 1983).

Eine Arbeit von Marsh et al. (2005b) betrachtet schließlich neben anderen auch den Zusammenhang zwischen Interesse im Fach Mathematik und der Schulnote in Mathematik in einem kreuzverzögerten Design (*cross-lagged panel* Design, CLP; vgl. Kapitel 9.3) mit zwei Messzeitpunkten (Schüler der 7. Klasse). Es finden sich zwar kleine, jedoch signifikante Effekte von 0.15 (T1 Interesse auf T2 Note) und 0.03 (T1 Note auf T2 Interesse). Aufgrund der deutlich unterschiedlichen Höhe der Pfadschätzungen kann auf Basis dieser Ergebnisse, zumindest für diesen Altersbereich, dementsprechend nicht von einer reziproken Beeinflussung ausgegangen werden.

Abschließend ist noch die relative Bedeutung von IM im Vergleich zum FSK bei der Vorhersage schulischen Erfolgs zu bewerten. Hier rücken insbesondere Studien, in denen beide Motivationskonstrukte gemeinsam untersucht wurden, in den Mittelpunkt. Diese belegen übereinstimmend einen höheren Anteil an aufgeklärter Varianz für FSK im Vergleich zur IM (z. B. 31% vs. 14%, B. Spinath et al., 2006; siehe z. B. auch Toussaint, 2007). In einigen Studien findet sich zudem für IM kein signifikanter Erklärungsbeitrag mehr, wenn das akademische Selbstkonzept mit in die Analyse aufgenommen wird (Bouffard et al., 2003; Marsh et al., 2005b; E. M. Skaalvik & S. Skaalvik, 2004).

Eine Betrachtung der Vorhersagebeiträge lediglich einzelner Faktoren, wie in der Mehrzahl der bisher berichteten Studien, greift allerdings zu kurz, um die Komplexität der Zusammenhänge adäquat abzubilden. Daher werden im nächsten Abschnitt solche Studien zusammenfassend dargestellt, die motivationale und kognitive Faktoren gemeinsam betrachten.

3.4.3 Relative Bedeutung motivationaler Faktoren unter Berücksichtigung allgemeiner kognitiver Fähigkeiten

Motivationale Variablen erhalten im Kontext der Vorhersage von Schulerfolg trotz ihrer, im Vergleich zu kognitiven Variablen, geringeren Vorhersagekraft vor allem aufgrund der ihnen zugesprochenen leichten Beeinflussbarkeit praktische Relevanz (B. Spinath et al., 2006). Studien, die sowohl motivationale als auch kognitive Faktoren gemeinsam untersuchen bestätigen zum einen Intelligenz als varianzstärksten Prädiktor, zeigen jedoch auch einen von Intelligenz unabhängigen Vorhersagebeitrag unterschiedlicher motivationaler Faktoren zwischen 4% und 9%, sowohl im Grundschul- als auch im Sekundarschulalter (z. B. Gose, Wooden & Muller, 1980; Helmke, 1992; Lloyd & Barenblatt, 1984; Schicke & Fagan, 1994; Steinmayr & B. Spinath, 2009).

Bei der Beurteilung der prädiktiven Validität von Motivation vor dem Hintergrund vorliegender Forschungsbefunde darf jedoch nicht unberücksichtigt bleiben, dass eine bloße Betrachtung ihres inkrementellen Erklärungsbeitrages eventuell zu einer Unterschätzung der Wichtigkeit motivationaler Faktoren führt (Grube & Hasselhorn, 2006; Helmke, 1992; B. Spinath & F. M. Spinath, 2005a). In diesem Fall bliebe unberücksichtigt, dass Motivations- und Intelligenzvariablen ebenfalls interkorrelieren und damit einen substanziellen gemeinsamen Varianzanteil aufweisen (Helmke, 1997; Sauer & Gamsjäger, 1996; B. Spinath et al., 2006). Dieser Konfundierung wird in den üblicherweise durchgeführten Korrelations- und Regressionsanalysen allerdings keine Rechnung getragen (Cooley & Lohnes, 1976; Helmke, 1992; B. Spinath & F. M. Spinath, 2005b). In einer Regressionsanalyse wird der gemeinsame Varianzanteil dem stärkeren Prädiktor (in diesem Fall der Intelligenz) zugerechnet, was folglich zu einer Unterschätzung des schwächeren Prädiktors (in diesem Fall der Motivation) führt. Zuverlässige Aussagen bezüglich der Vorhersage schulischen Erfolgs durch allgemeine kognitive Fähigkeiten und Motivation lassen dementsprechend nur solche Studien zu, die die gegenseitige Abhängigkeit der Prädiktoren berücksichtigen (Heller & A. Ziegler, 1998).

Um die relative Bedeutung motivationaler Variablen im Vergleich zu anderen Einflussfaktoren adäquat beurteilen zu können, wurde in einigen Studien eine *Kommunalitätenanalyse* durchgeführt (Cooley & Lohnes, 1976). Durch dieses Verfahren kann die

vorherzusagende Kriteriumsvarianz in spezifische und konfundierte Erklärungsbeiträge der Prädiktoren zerlegt werden (Cohen, Cohen, West & Aiken, 2003; B. Spinath et al., 2006). Es sei hier allerdings darauf hingewiesen, dass bisher nur wenige Studien vorliegen, die mit Hilfe dieses Verfahrens spezifische und gemeinsame Varianzanteile differenziert betrachten (Helmke, 1992; Helmke, 1997; Helmke & Weinert, 1997). Zudem ist auch dieses Vorgehen mit einigen methodischen Mängeln behaftet (für eine Diskussion siehe Schneider, 1991), weshalb in aktuellen Studien zur Beurteilung der prädiktiven Bedeutsamkeit bei simultan betrachteten und miteinander korrelierenden Einflussfaktoren auf die Verwendung von Strukturgleichungsmodellen zurückgegriffen wird. Diese bieten die Möglichkeit der Modellierung von Messfehlern der gemessenen Variablen und der Berücksichtigung spezifischer korrelativer Beziehungen zwischen den betrachteten Prädiktoren (Weiber & Mühlhaus, 2010).

Unter Verwendung einer Kommunalitätenanalyse berichten beispielsweise Sauer und Gattringer (1986) bei einem gemeinsamen Vorhersagebeitrag schulischen Erfolgs (Schulnoten) durch Intelligenz und Motivationsmerkmale von ca. 49%, einen spezifischen Anteil von ca. 31% durch kognitive Fähigkeiten und einen spezifischen Anteil von ca. 5% für Motivation. Der konfundierte Varianzanteil lag bei rund 13%. Bei Helmke (1992) lag der konfundierte Varianzanteil im Fach Mathematik mit 30% noch deutlich höher. Von besonderer Relevanz für die vorliegende Studie sind aufgrund der korrespondierenden Operationalisierung kognitiver Fähigkeiten sowie des FSK und der IM Ergebnisse auf Basis von Daten der englischen Zwillingsstudie TEDS. Die durch Lehrer eingeschätzte akademische Leistung basiert hier auf den Kriterien des sogenannten *UK National Curriculum*⁵ (NC).

B. Spinath et al. (2006) konnten in dieser Stichprobe (9-jährige Grundschul Kinder; N=1.678) durch eine Kommunalitätenanalyse ebenfalls den *g-Faktor* als stärksten Prädiktor sowohl bei gemeinsamer Betrachtung mit dem FSK als auch der IM bestätigen. Für beide Motivationsmaße fand sich ein spezifischer Varianzanteil, der für das FSK jedoch deutlich höher ausfiel (FSK: zwischen 25% in Mathematik und 31% in Englisch; IM: zwischen 12% in Englisch und 14% in Mathematik). Als entscheidender Befund dieser Studie kann der demonstrierte, substanziell konfundierte Varianzanteil der Prädiktoren bewertet werden. Dieser betrug für das FSK zwischen 10% (Naturwissenschaft) und 24% (Englisch), sowie zwischen 7% (Mathematik) und 8% (Englisch) in der IM (B. Spinath et al., 2006).

Auch auf Grundlage von Daten aus der ersten Erhebungswelle der KoSMoS Studie

⁵ Diese Kriterien stellen einen einheitlichen Bewertungsschlüssel für die Benotung von Kindern bereit. Es werden fünf Stufen der Leistung für jedes Schulfach unterschieden, jede davon umfasst einen breiten Bereich von Fähigkeiten (siehe Walker et al., 2004 für eine umfassende Beschreibung).

liegen Ergebnisse einer Kommunalitätenanalyse vor (Toussaint, 2007). Die Bedeutung motivationaler Faktoren zur Vorhersage von schulischem Erfolg operationalisiert durch Fachnoten wurde wiederum bestätigt. Allerdings lag in dieser Studie keine prädiktive Dominanz allgemeiner kognitiver Fähigkeiten vor (Toussaint, 2007). Beispielsweise lässt sich die im Schulfach Deutsch insgesamt durch FSK und kognitive Fähigkeit aufgeklärte Varianz von 53% zu 4% auf g , zu 79% auf FSK und zu 17% auf einen gemeinsamen Varianzanteil zurückführen.

Dieser Befund kann zum einen durch die eingesetzte Heimtestung kognitiver Fähigkeiten durch die Eltern, die in ihrer psychometrischen Güte wohl hinter einer face-to-face Testung zurückbleibt (Pooch, 2006), erklärt werden (Toussaint, 2007). Zum anderen kann hier auch die Operationalisierung des Kriteriums durch Schulnoten eine Rolle spielen. Wie unter Abschnitt 3.4.1 und 3.4.2 erläutert, fallen die Korrelationen zwischen Motivation und schulischer Leistung besonders dann hoch aus, wenn, wie bei Toussaint (2007), die Erfassung der Motivation domänenspezifisch erfolgt und mit den entsprechenden Fachnoten in Zusammenhang gebracht wird. Bei B. Spinath et al. (2006) wurden im Gegensatz dazu Lehrerratings herangezogen, die an Hand eines nationalen Curriculums den Leistungsstand ihrer Schüler einschätzen sollten (vgl. hierzu Walker et al., 2004). Einen weiteren Hinweis auf die Höhe der prädiktiven Validität in Abhängigkeit der Operationalisierung des schulischen Erfolgs durch Noten liefert eine Studie von Steinmayr und B. Spinath (2009). Hier fanden sich für den Sekundarschulbereich ebenfalls höhere Beiträge zur Varianzaufklärung durch das FSK und *Werte* als durch die jeweiligen Intelligenzfaktoren (verbal bzw. numerisch).

Helmke und Schrader (2006) stellen zusammenfassend fest, dass Ergebnisse aus Studien unter Verwendung der Kommunalitätenanalyse die prädiktive Dominanz kognitiver Fähigkeiten gegenüber anderen Determinanten der Schulleistung bestätigen. Dennoch zeigt sich, dass ein großer Teil der Kriteriumsvarianz nur durch kognitive und motivationale Variablen gemeinsam erklärt werden kann (vgl. hierzu auch Helmke, 1997). Weiterhin lassen Ergebnisse auf Grundlage von Kommunalitätenanalysen keinerlei Aussagen über die Richtung der wechselseitigen Beeinflussung zu (B. Spinath & F. M. Spinath, 2005a). Diese können nur auf Grundlage längsschnittlicher Datenanalysen getroffen werden.

Eine längsschnittliche Erweiterung der Studie von B. Spinath et al. (2006) auf das FSK und die akademische Leistung im Alter von 12 Jahren nehmen Chamorro-Premuzic, Harlaar, Greven und Plomin (2010) vor. Hier wurde im Unterschied zu B. Spinath et al. (2006) allerdings eine gepoolte FSK Schätzung sowie ein ebenfalls gepoolter Wert des NC aus Englisch, Mathematik und Naturwissenschaft gewählt. Die Analyse erfolgte hier nicht anhand einer Kommunalitätenanalyse sondern unter Rückgriff auf eine Strukturgleichungsmodellierung. Unter Berücksichtigung einer reziproken Bezie-

hung beider Variablen (z. B. Marsh & Craven, 2006) wurde ein CLP Modell, in dem aufgrund deren positiver Relation die Intelligenzmessung für beide Konstrukte kontrolliert wurde, getestet.⁶ Die Ergebnisse legen nahe, dass FSK und akademische Leistung im Alter 9, auch unabhängig von Intelligenz, sowohl die Fähigkeitseinschätzung als auch die akademische Leistung im Alter 12 beeinflussen. Die Pfadschätzungen beider Kreuzpfade waren dabei vergleichbar hoch ($\beta = 0.13$ von Leistung $Z1$ auf FSK $Z2$; $\beta = 0.11$ von FSK $Z1$ auf Leistung $Z2$), was zusätzlich eine reziproke Beeinflussung auch unter Kontrolle der Intelligenz nahelegt.

3.5 Individuelle Entwicklungsstabilität motivationaler Variablen

Ergänzend zur Betrachtung der prognostischen Güte motivationaler Variablen findet in der pädagogischen Psychologie vor allem die Frage nach der Entwicklung motivationaler Personenmerkmale im Verlauf der Schulzeit Beachtung. Die hohe Relevanz, die dieser Frage beigemessen wird ergibt sich aus der Annahme heraus, dass eine beständig hohe Motivation, sich neuen Herausforderungen zu stellen und die Bereitschaft, lebenslang zu lernen, wünschenswerte und unerlässliche Eigenschaften für einen erfolgreichen schulischen und beruflichen Werdegang darstellen (Filipp, 2006).

Während ihrer Schulzeit werden Kinder und Jugendliche von Seiten der Lehrer, Mitschüler und auch der Eltern kontinuierlich mit expliziten und impliziten Rückmeldungen bezüglich ihrer Leistungsfähigkeit konfrontiert. Diese Rückmeldungen werden von den Kindern beispielsweise in die Selbstwahrnehmung der eigenen Kompetenzen integriert (Filipp, 2006). Sowohl Eltern als auch Lehrern kommt in diesem Prozess eine wichtige Rolle als Informationsquelle zu, wobei der Einfluss der Lehrer über die Schullaufbahn hinweg zu, der Einfluss der Eltern hingegen abnimmt (B. Spinath & F. M. Spinath, 2005a).

Innerhalb der psychologischen Forschung wurde die Frage nach der Entwicklung motivationaler Personenmerkmale bislang insbesondere aus zwei Perspektiven heraus betrachtet: Zum einen wurden speziell das *Leistungsmotiv* und das *Interesse* unter ontogenetischen Gesichtspunkten untersucht (Holodynski, 2007; Schiefele, 2009), zum anderen wurden fast alle motivationalen Merkmale hinsichtlich ihrer quantitativen Verläufe analysiert (Schiefele, 2009).

Nachfolgend wird zunächst die umfangreiche Forschung zur Entwicklung von Selbstkon-

⁶ Die Analyse erfolgte mit den Residuen der latenten Faktorwerte (Chamorro-Premuzic et al., 2010).

zepten und Einflüssen auf die Höhe von Selbstkonzepten zusammenfassend dargestellt. Anschließend werden Befunde zu Entwicklungsverläufen von intrinsischer Motivation und Interesse betrachtet.

3.5.1 Individuelle Entwicklungsstabilität des Fähigkeitsselbstkonzeptes

In der Literatur findet sich weithin Bestätigung für eine statistisch signifikante Mittelwertsveränderung des Selbstkonzeptes über die Schullaufbahn hinweg, wobei sich ein deutlicher Abwärtstrend abzeichnet. Während Kinder im frühen Grundschulalter ihre eigenen Fähigkeiten und Kompetenzen in einem sehr positiven Licht sehen (vgl. D. Stipek & Iver, 1989), wird das Selbstbild der Kinder bereits im Verlauf ihrer Grundschulzeit zunehmend realistischer und entwickelt sich unter Umständen sogar zu einer pessimistischen Selbstwahrnehmung (z. B. Dweck, 2002; Eccles et al., 1993b; Guay et al., 2003; Jacobs, Lanza, Osgood, Eccles & Wigfield, 2002; B. Spinath & F. M. Spinath, 2005b; B. Spinath & Steinmayr, 2008; Wigfield et al., 1997). Zudem bildet sich bereits im frühen Grundschulalter eine immer differenziertere Wahrnehmung der eigenen Fähigkeit hinsichtlich einzelner Schulfächer aus, die im Laufe der Entwicklung noch weiter zunimmt (Eccles et al., 1993b; Marsh, Byrne & Shavelson, 1992; Wigfield et al., 1997). So fanden Eccles et al. (1993b) beispielsweise ein Absinken der Fähigkeitsselbstkonzepte bei Grundschulern im Alter zwischen 7 und 10 Jahren für die Unterrichtsfächer Lesen, Mathematik und Musik, wobei sich dieser Trend häufig bis in die Sekundarstufe fortsetzt (Wigfield & Eccles, 2000). Auch B. Spinath und Steinmayr (2008) finden sinkende Fähigkeitsselbsteinschätzungen bei Grundschulern über vier Messzeitpunkte hinweg. Durch die Analyse latenter Wachstumskurvenmodelle konnten zudem interindividuelle Unterschiede in den Verlaufskurven aufgezeigt werden (B. Spinath & Steinmayr, 2008).

In Bezug auf die interindividuelle Stabilität weisen Befunde aus verschiedenen Studien auf eine bereits im Grundschulalter nachweisbare, hohe interindividuelle Stabilität schulbezogener Selbstkonzepte hin (z. B. Helmke, Schrader & Lehneis-Klepper, 1991, B. Spinath & F. M. Spinath, 2005b, Marsh, Craven & Debus, 1998, Wigfield et al., 1997). Guay et al. (2003) berichten für Schüler in der zweiten, dritten und vierten Klasse, die im Abstand von jeweils einem Jahr insgesamt dreimal befragt wurden, eine mittlere Stabilität von 0.43 für die jüngste Kohorte und von 0.65 für die älteste Kohorte. Selbst bei einem Wechsel der Bezugsgruppe, z. B. nach dem Ende der Schulzeit, finden sich Stabilitätskoeffizienten in ähnlicher Höhe (Marsh, Trautwein, Lüdtke, Baumert & Köller, 2007).

Wie lässt sich nun dieses zum Teil rapide Absinken von Fähigkeitsüberzeugungen vor al-

lem bei Kindern zwischen 6 und 10 Jahren erklären? Hypothesen über die Entwicklung des Selbstkonzeptes und über Bedingungen, die die Höhe des Selbstkonzeptes bestimmen, haben in viele Bereiche psychologischer Forschung Eingang gefunden. Aus einer entwicklungspsychologischen Perspektive heraus beschreibt Susan Harter die Struktur und die zentralen Inhalte von Selbstkonzepten sowie deren Übereinstimmung mit der Wirklichkeit für sechs Altersstufen vom Säuglingsalter bis zum späten Jugendalter (Harter, 1998, 1999). Hiernach sind die Selbstkonzepte bis weit in die Kindheit hinein stark positiv gefärbt. In Folge einer zunehmenden Integration von externen, auch negativen Rückmeldungen über die eigenen Fähigkeiten steigt die Genauigkeit der Selbsteinschätzungen im Laufe der Zeit immer mehr an. Hierdurch gelingt eine zunehmende Ausdifferenzierung des eigenen Selbstkonzeptes, das schließlich am Ende der Jugendzeit relativ stabile Überzeugungen und Werte reflektiert (Harter, 1999). Dieser Annahme folgend ist es wahrscheinlich, dass es Kindern mit zunehmendem Alter immer besser gelingt, die Bewertungen der eigenen Leistungen durch das soziale Umfeld – im schulischen Kontext geschieht dies vor allem durch die Lehrer (B. Spinath & F. M. Spinath, 2005a) – präziser zu interpretieren und in das Fähigkeitsselbstbild zu integrieren (Wigfield & Eccles, 2002). Die Bewertung der eigenen Leistungen (z. B. durch Schulnoten) und die Gelegenheiten zu sozialen Vergleichen mit den Leistungen gleichaltriger Kinder werden mit dem Schuleintritt automatisch häufiger (Dweck, 2002; B. Spinath & F. M. Spinath, 2005b; Wigfield & Eccles, 2002), was den Beginn des Abwärtstrends von Fähigkeitsselbsteinschätzungen gerade im Alter zwischen 6 und 10 Jahren plausibel macht. Zudem lernen Kinder im Laufe ihrer Schulzeit zwischen unterschiedlichen Schulfächern und ihren Fähigkeiten in diesen Fächern besser zu differenzieren (B. Spinath & Steinmayr, 2008). Somit können Schüler mit zunehmendem Alter ihre Einschätzungen bezüglich der eigenen Fähigkeit immer genauer und realistischer vornehmen, was meist in weniger positiven schulischen Selbstkonzepten resultiert (Dweck, 2002; Guay et al., 2003; Wigfield & Eccles, 2002).

Neben der zuvor beschriebenen Perspektive wurden in der Literatur vor allem zwei Phänomene diskutiert, die Aufschluss über die Höhe der fachspezifischen Selbstkonzepte geben. Beide können als *Bezugsrahmeneffekte* bezeichnet werden. Sowohl das sogenannte *Internal/External-Frame-of-Reference-Modell* (I/E-Modell; Marsh, 1986) als auch der *Big-Fish-Little-Pond-Effekt* (BFLPE; Marsh, 1987) thematisieren die Auswirkungen von Leistungsvergleichen auf Selbstkonzepte. Im *I/E-Modell* wird angenommen, dass Personen sowohl externe (d. h. soziale) als auch interne (d. h. dimensionale) Vergleiche zur Bildung ihres FSKs nutzen, während mit dem BFLPE konkretisiert wird, auf welche Weise der externe Bezugsrahmen das FSK von Schülern beeinflusst (Dickhäuser & Galfe, 2004; Lüdtke & Köller, 2002; Möller & Trautwein, 2009).

Das *I/E-Modell* von Marsh (1986) gilt als empirisch sehr gut belegtes Modell um zu

erklären, warum das mathematische und das verbale Selbstkonzept nicht oder nur gering miteinander korrelieren, obwohl die Leistungen in diesen Fächern normalerweise substanzielle Zusammenhänge aufweisen (vgl. Abschnitt 3.4.1). In diesem Modell wird angenommen, dass fachspezifische Selbstkonzepte durch internale und externale Vergleiche beeinflusst werden (E. M. Skaalvik & S. Skaalvik, 2004). Schüler benutzen zunächst einen externalen Bezugsrahmen zur Beurteilung der eigenen Leistung, indem sie sich mit der wahrgenommenen Leistung anderer Schüler vergleichen und diesen relativen Eindruck als eine Quelle ihres akademischen Selbstkonzeptes nutzen (Marsh, Byrne & Shavelson, 1988). Diese interindividuellen (sozialen) Vergleiche führen bei Schülern mit guten Leistungen zu einem hohen Selbstkonzept in diesem Fach und umgekehrt bei Schülern mit weniger guten Leistungen zu schwachen Selbstkonzepten. Statistisch ergibt sich daraus eine positive Korrelation zwischen Leistung und Selbstkonzept innerhalb eines Schulfaches (Möller & Trautwein, 2009). Dieses Phänomen wurde in einer Reihe von Studien belegt (vgl. Abschnitt 3.4.1). Schüler wenden neben diesen externalen Vergleichen aber auch internale Vergleiche an, d. h. sie vergleichen ihre eigenen Leistungen in einem Fach mit ihren Leistungen in einem anderen Fach und zwar unabhängig von externalen Vergleichen (Marsh, Parker & Barnes, 1985). Diese intraindividuellen bzw. dimensionalen Vergleiche führen dazu, dass beispielsweise ein Schüler, der seine Fähigkeiten in Mathematik schlechter einschätzt als die seiner Klassenkameraden, dennoch ein positives mathematisches Selbstkonzept entwickelt, weil er seine, in Relation zu anderen Fächern, bessere Fähigkeit in Mathematik sieht (Marsh et al., 1985; Möller & Trautwein, 2009). Rein rechnerisch sollten internale Vergleiche dementsprechend zu negativen Korrelationen zwischen verschiedenen akademischen Selbstkonzepten führen, weil die Leistungen in den verschiedenen Schulfächern miteinander verglichen werden (E. M. Skaalvik & S. Skaalvik, 2004). Beide Vorhersagen wurden in einer Reihe von Studien bestätigt (z. B. B. M. Byrne & Shavelson, 1987; Lüdtke & Köller, 2002; Marsh, 1986; Marsh et al., 1988, 1991; E. M. Skaalvik & Rankin, 1995). In einer aktuellen Metaanalyse integrieren Möller und Trautwein (2009) die Daten von 69 Studien in denen die postulierten Zusammenhänge des I/E-Modells untersucht wurden. Die Vorhersagen des Modells konnten insgesamt bestätigt werden, interessanterweise zeigte sich das Zusammenhangsmuster theoriekonform relativ unabhängig vom Alter der Schüler und sowohl für die Beurteilung schulischer Leistung an Hand von Schulnoten als auch für Ergebnisse aus objektiven Leistungstest (Möller & Trautwein, 2009).

Mit dem BFLPE wird der negative Effekt leistungsstarker Referenzgruppen auf das Selbstkonzept von Schülern und Studierenden beschrieben (Marsh, 1987). Der Bezugsrahmen für die im I/E Modell angenommenen externalen (sozialen) Vergleiche wird hiernach primär durch die Schulklasse definiert, der ein Schüler angehört (Möller & Trautwein, 2009). Die Existenz des BFLPEs ist mittlerweile empirisch gut belegt (Trautwein & Lüdtke, 2005). Arbeiten zum BFLPE weisen darauf hin, dass Schüler mit

einer bestimmten Leistungsstärke genau dann ein relativ hoch ausgeprägtes Selbstkonzept aufweisen, wenn sie sich in leistungsschwachen Klassen befinden. Hingegen haben Schüler mit identischer Leistungsstärke niedrigere schulische Selbstkonzepte, wenn sie in leistungsstarken Klassen platziert werden (z. B., Dickhäuser & Galfe, 2004; Schwarzer, Lange & Jerusalem, 1982; Trautwein, Lüdtke, Köller & Baumert, 2006a; für einen Überblick siehe, Köller, 2004). Erklärt wird dieser Effekt dadurch, dass es in leistungsstarken Klassen mehr Möglichkeiten für soziale Aufwärtsvergleiche mit leistungsstärkeren Mitschülern gibt. Diese Aufwärtsvergleiche haben dann negative Konsequenzen für die selbst eingeschätzten Fähigkeiten (Möller & Trautwein, 2009). Besonders gut kann der BFLPE beim Übergang von der Grundschule in die weiterführende Schule beobachtet werden und fällt besonders gravierend bei Gymnasialschülern aus, denn diese Schulform bietet in der Regel besonders häufig Gelegenheit zu sozialen Aufwärtsvergleichen. Stärkere Bezugsrahmeneffekte finden sich dann, wenn in den verwendeten Selbstkonzeptmaßen implizit oder explizit nach sozialen Vergleichen gefragt wird (z. B. Noten oder Leistungen), im Gegensatz zu Instrumenten bei denen z. B. ein kriterialer Vergleichsmaßstab verwendet wird (siehe Filipp, 2006; Möller & Trautwein, 2009). Als praktische Implikation kann nun allerdings nicht geschlossen werden, dass Eltern ihre Kinder gezielt in leistungsschwache Klassen platzieren sollten, denn im Hinblick auf die Leistungsentwicklung finden sich Hinweise auf ein gegenteiliges Muster: Hier profitieren die Kinder in der Regel von der Zugehörigkeit zu einer leistungsstarken Gruppe (Becker, Lüdtke, Trautwein & Baumert, 2006).

Der BFLPE ist jedoch nicht alleine auf das Selbstkonzept beschränkt. In jüngerer Zeit konnte an Hand von Daten der *Third International Mathematics and Science Study* TIMSS sowie der PISA-Studie neben den erwarteten Effekten auf das mathematische Selbstkonzept auch Effekte auf das Interesse an Mathematik nachgewiesen werden (Köller, Daniels, Schnabel & Baumert, 2000; Trautwein, Lüdtke, Marsh, Köller & Baumert, 2006b).

3.5.2 Individuelle Entwicklungsstabilität intrinsischer Motivation

Werden die Ergebnisse zahlreicher empirischer Studien zusammenfassend betrachtet, findet sich analog zu den zuvor dargestellten Befunden zum FSK, auch für verschiedene Indikatoren der Lernmotivation in verschiedenen Altersgruppen bzw. Schulstufen eine generelle Abnahme der Lernmotivation. Dieser Abwärtstrend beginnt mit den ersten Schuljahren und setzt sich kontinuierlich bis zum Ende der Schulzeit fort.

Auch das Interesse an bestimmten Schulfächern, bzw. der Schule allgemein, folgt diesem Abwärtstrend, dieser scheint aber insbesondere bei naturwissenschaftlichen Fächern besonders ausgeprägt zu sein (z. B. Baumert & Köller, 1998; Krapp, 2006; Schiefele,

2009). In gleicher Weise verhält es sich in Bezug auf die Einstellung gegenüber der Schule, den aufgabenbezogenen Wertüberzeugungen (Nützlichkeit, Wichtigkeit) und Indikatoren intrinsischer Motivation (z. B. Bouffard et al., 2003; Gottfried, Fleming & Gottfried, 2001; Gottfried, Marcoulides, Gottfried, Oliver & Guerin, 2007; Jacobs et al., 2002; B. Spinath & Steinmayr, 2008; Wigfield et al., 1997), sowohl für die Grundschule als auch in der weiterführenden Schule. B. Spinath und F. M. Spinath (2005b) finden beispielsweise ein Absinken der Lernmotivation sowie der Kompetenzüberzeugungen von Grundschulkindern (Effektstärke zwischen 0.04 und 0.34) über ein Intervall von zwei Jahren hinweg (in dieser Zeit wurden die Kinder halbjährlich befragt) in allen fünf untersuchten Kohorten. Hierbei war das Absinken der Lernmotivation in den älteren Kohorten besonders stark ausgeprägt (B. Spinath & F. M. Spinath, 2005b). Auch Harter (1981) konnte eine Verringerung verschiedener Komponenten der IM von der 3. bis zur 9.Klasse aufzeigen.

Der erläuterte negative Verlauf der IM lässt jedoch keinen Schluss über die interindividuelle Stabilität dieses Merkmals zu. Diesbezüglich kann eine hohe Stabilität der IM über die gesamte Schullaufbahn hinweg als empirisch gesichert gelten, die Rangreihe der Schüler bleibt also im Wesentlichen erhalten. Die berichteten Stabilitätskoeffizienten liegen üblicherweise in einem mittleren Bereich für Kinder im Grundschulalter (Bouffard et al., 2003; B. Spinath & F. M. Spinath, 2005b; B. Spinath & Steinmayr, 2008; Wigfield et al., 1997) und in einem mittleren bis hohen Bereich für Kinder im Sekundarschulalter (Gottfried et al., 2001; Helmke, 1993; Marsh et al., 2005b). Die interindividuelle Stabilität nimmt also über den Verlauf der Schulzeit hinweg zu.

Obwohl es hinreichende Belege für ein generelles Absinken intrinsischer Lernmotivation während der Schulzeit gibt, besteht noch immer kein klarer Konsens über die Ursachen dieser Entwicklung (B. Spinath & F. M. Spinath, 2005b). Eine häufig angestellte Vermutung diesbezüglich ist, dass die Entwicklung von intrinsischen Werten und Kompetenzüberzeugungen von Kindern parallel verläuft. Kinder sollten dieser Annahme zu Folge eine besonders hohe Lernmotivation zeigen, wenn sie ihre eigene Kompetenz in diesem Bereich als hoch einschätzen (Eccles & Wigfield, 1995; Eccles, 2005). Wird nun davon ausgegangen, dass eine zu Beginn der Schulzeit sehr optimistische Sicht der eigenen Fähigkeiten durch eine zunehmend realistische Selbsteinschätzung der Kinder ersetzt wird (siehe z. B. Newman, 1984; D. J. Stipek & D. H. Daniels, 1988; Wigfield et al., 1997), würde ein Absinken der Fähigkeitsselbstüberzeugungen in einem Bereich gleichzeitig mit einem Absinken der intrinsischen Lernmotivation in korrespondierendem Bereich führen (B. Spinath & F. M. Spinath, 2005b). Empirisch wurde diese Frage vor allem im Rahmen des Erwartungs-Wert-Modells von Eccles (Eccles et al., 1983) untersucht. Zwar konnte in Studien eine über die Zeit hinweg stärker werdende korrelative Beziehung zwischen intrinsischen Werten und Kompetenzüberzeugungen nachgewiesen werden (Eccles & Wigfield, 1995; Obach, 2003; Wigfield et al., 1997),

eine kausale Beziehung zwischen beiden Konstrukten konnte hingegen in der großen Mehrheit an Untersuchungen nicht bestätigt werden (Obach, 2003; Nurmi & Aunola, 2005; E. M. Skaalvik & Valas, 1999; B. Spinath & F. M. Spinath, 2005b; B. Spinath & Steinmayr, 2008). Lediglich Jacobs et al. (2002) schließen aus den Ergebnissen ihrer Analyse, dass Kompetenzüberzeugungen einen substanziellen Varianzanteil in der Veränderung von Werten über die Zeit hinweg erklären (für eine detaillierte Diskussion dieser widersprüchlichen Befunde siehe Eccles, 2005).

Bezugnehmend auf die verfügbare empirische Evidenz nimmt Schiefele (2009) drei unterschiedliche Prozesse des Absinkens schulischer Interessen und damit letztlich auch der intrinsischen Motivation an. Erstens stehen bestimmte Unterrichtsmerkmale, z. B. Vernachlässigung der Alltagserfahrungen und der Interessen der Schüler, sowie der restriktive, wenig Raum für Selbstbestimmung bietende Charakter schulischer Lernumwelten, der Entfaltung schulfachbezogener Interessen entgegen. Zweitens entwickeln die Schüler mit zunehmendem Alter auch stabile außerschulische Interessen, die dann in Konkurrenz zur Schule treten. Drittens führt die Wahrnehmung eigener Fähigkeiten im Laufe der Schulzeit zu eigenen Interessenschwerpunkten, die wiederum die Aufgabe oder Abwertung anderer Interessensbereiche bedingen. Insbesondere der letztgenannte Punkt kann durchaus auch als positiver Differenzierungsprozess einer fähigkeitsabhängigen Spezialisierung interpretiert werden (Schiefele, 2009).

3.6 Abschließende Bewertung

Zusammenfassend betrachtet bestätigen die Ergebnisse der zuvor dargestellten Studien insgesamt die vor allem in der Pädagogischen Psychologie postulierte Bedeutung motivationaler Grundvoraussetzungen für die Vorhersage schulischen Erfolgs, sowohl im Grundschul- als auch im Sekundarschulbereich. Der gefundene positive Zusammenhang bleibt auch nach Kontrolle des stärksten Einzelprädiktors *Intelligenz* (vgl. Kapitel 2.3) bestehen. In methodisch aufwändigen Studien wurde zudem eine reziproke Beeinflussungsrichtung von FSK und akademischer Leistung weitgehend bestätigt. Generell zeichnet sich jedoch ein negativer Trend (im Sinne intraindividuelle Stabilität) sowohl in der Fähigkeitsselbsteinschätzung als auch der IM bzw. des Interesses über die Schullaufbahn hinweg ab. Es ist zu vermuten, dass sich ein besonders ausgeprägter negativer Trend unmittelbar nach Übergangserfahrungen, wie sie z. B. nach dem Übertritt von der Grundschule in die Sekundarschule auftreten, und dem damit verbundenen Wechsel der Bezugsgruppe findet. Auf interindividueller Ebene bleiben motivationale Voraussetzungen hingegen auf einem vergleichbar hohen, mittleren Niveau bestehen.

Generierte Erklärungsmodelle dieser Entwicklung beziehen sich in erster Linie auf Bezugsrahmeneffekte und sind vor allem für das FSK empirisch belegt.

Gleichwohl ist die Generalisierbarkeit der Ergebnisse mit gewissen Einschränkungen verbunden, denn die Höhe der berichteten Zusammenhänge unterliegt in Abhängigkeit der gewählten Operationalisierung der Prädiktoren (z. B. fachspezifisch oder global) sowie des Kriteriums (Fachnoten, Einschätzung der Lehrer anhand festgesetzter Kriterien, Schulleistungstests) teilweise beträchtlichen Schwankungen. Ferner liegen den Analysen zum Teil recht kleine Stichproben zugrunde und die Mehrzahl der betrachteten Stichproben entstammt dem anglo-amerikanischen Raum und bezieht sich somit auf eine vom deutschen Schulsystem sehr abweichende Schullaufbahn.

Darüber hinaus ist auch auf methodische Mängel in den gewählten Analysemethoden einiger Studien hinzuweisen. So stellt die häufig verwendete multiple Regressionsanalyse zwar prinzipiell eine geeignete Methode zur Analyse der Vorhersage individueller Unterschiede in einer Domäne durch individuelle Unterschiede in einer anderen Domäne dar, sie geht jedoch auch mit gravierenden Problemen einher, die zu verzerrten Schätzungen führen können (Schneider, 1991). Zum einen wird in den Modellannahmen lediglich das Kriterium als Zufallsvariable deklariert, während für die Prädiktoren eine fehlerfreie Messung angenommen wird. Eine solche fehlerfreie Messung ist allerdings in der Regel nicht erfüllt, was sich wiederum in verzerrten Varianzschätzungen der Prädiktoren und damit fehlerbehafteten multiplen Korrelationskoeffizienten niederschlägt (Schneider, 1991). Eine weitere Schwierigkeit besteht in der bereits erwähnten, häufig vorliegenden Interkorrelation der Prädiktoren, wodurch die spezifischen Anteile von Prädiktoren an der Kriteriumsvarianz nicht mehr exakt zu bestimmen sind. Allgemein gilt, dass die Verzerrung der Schätzwerte aus der multiplen Regression um so gravierender ist, je geringer die Reliabilität der Prädiktoren und je höher deren Interkorrelation ausfällt (Schneider, 1991).

Um diese Probleme zu umgehen, wird in einigen Studien auf das Verfahren der Kommunalitätenanalyse zurückgegriffen. Aber auch bei Anwendung dieser Methode kommt es zu Interpretationsproblemen, insbesondere, wenn mehr als zwei Prädiktoren betrachtet werden (zu dieser Problematik siehe auch Pedhazur, 1982; Schneider, 1991).

Durch die Anwendung latenter Strukturgleichungsmodelle können diese methodischen Schwierigkeiten vermieden werden. Im Rahmen dieser Verfahren können sowohl Messfehler als auch Interkorrelationen zwischen den Prädiktoren berücksichtigt werden, wodurch eine exaktere Bestimmung der Varianzbeiträge einzelner Prädiktoren möglich wird (vgl. Kapitel 9).

Insgesamt kann demnach konstatiert werden, dass es speziell im deutschsprachigen Raum an Längsschnittstudien fehlt, die den Beitrag verschiedener motivationaler Variablen zur Vorhersage schulischen Erfolgs unter Rückgriff auf methodisch angemessene

Designs untersuchen. In diesem Zusammenhang hat sich vor allem das REM als Modell zur Beurteilung der längsschnittlichen Beeinflussung zwischen FSK und schulischer Leistung bewährt. Es mangelt bisher jedoch an Studien, die dieses Modell im Rahmen eines *Multi-Kohorten-Sequenz-Designs* testen. Auch ist die Befundlage bzgl. eines Vorherrschens des *skill-development Effektes* in niedrigeren Klassenstufen uneinheitlich und eine vergleichende Betrachtung verschiedener Schulfächer bei fachspezifischer Operationalisierung steht bis dato noch aus. Ferner würde die Übertragung des Modells auf weitere motivationale Variablen neben dem FSK eine interessante Erweiterung darstellen. Nicht zuletzt wurde in den Arbeiten von Marsh nicht um Intelligenz kontrolliert, obwohl diese den stärksten Einzelprädiktor schulischen Erfolgs darstellt. Hier wäre zu vermuten, dass die berichteten Zusammenhänge insgesamt etwas nach unten zu korrigieren sind.

Die gefundenen phänotypischen Beziehungen zwischen den Variablen stellen hier jedoch nur einen ersten Schritt im Verständnis der Zusammenhänge dar. Zur Aufklärung der Ursachen interindividueller Unterschiede in den Variablen muss auf genetisch sensitive Daten zurückgegriffen werden. Mit Hilfe dieser Designs ist es möglich, die beobachtete phänotypische Varianz in genetische und umweltbedingte Anteile aufzusplitten. Hierdurch ergeben sich beispielsweise wertvolle Zusatzinformationen für die Spezifizierung von Entwicklungsmodellen wie dem Erwartungs-Wert-Modell von Eccles und Kollegen (F. M. Spinath, Spinath & Plomin, 2008a). Daneben können die Ursachen der phänotypisch beobachteten interindividuellen Stabilität in motivationalen Variablen aufgedeckt werden. Durch verhaltensgenetische Analysen ist zudem die Untersuchung der genetischen Beeinflussung von *cross-lagged-Beziehungen* zwischen den Variablen möglich.

Durch die Verortung der vorliegenden Studie in der Zwillingsstudie KoSMoS wird die Beantwortung dieser noch offenen Forschungsfragen möglich. Aus diesem Grund werden im folgenden Kapitel zunächst die Grundlagen verhaltensgenetischer Forschung dargelegt. In einem zweiten Teil wird der vorhandene empirische Forschungsstand zu Ursachen interindividueller Unterschiede in schulischem Erfolg sowie der Prädiktoren Intelligenz und Motivation dargestellt. Die zur Beantwortung der Fragestellungen vorliegender Arbeit umgesetzten verhaltensgenetischen Analysen werden in Kapitel 10 vorgestellt.

4 Der Ansatz der Verhaltensgenetik zur Erklärung der Ursachen interindividueller Unterschiede

Die Verhaltensgenetik stellt einen Teilbereich der Genetik dar und beschäftigt sich mit der Frage, in welchem Ausmaß beobachtbare interindividuelle Unterschiede in einem Merkmal auf die genetische Veranlagung einerseits und auf die Umwelt andererseits zurückzuführen sind. Das Hauptziel verhaltensgenetischer Forschung liegt demnach in der Erklärung der Ursachen interindividueller Unterschiede. Sie verbindet hierzu Genetik und Psychologie, indem sie Methoden und Ergebnisse der Genetik auf die Erforschung von Verhalten anwendet (Brocke, Spinath & Strobel, 2004).

Initial für die Methode der heutigen Verhaltensgenetik war die Arbeit von Francis Galton, der bereits 1869 in seinem Werk „Hereditary Genius“ eine hohe Erblichkeit geistiger Fähigkeiten berichtete (Galton, 1869). Seither wurde eine Vielzahl an verhaltensgenetischen Studien durchgeführt und kaum ein anderes Gebiet innerhalb der empirischen Psychologie hat in jüngerer Vergangenheit eine vergleichbar schnelle, methodische Entwicklung erfahren.

Im Folgenden werden zunächst die Grundlagen der verhaltensgenetischen Forschung vorgestellt. Hierbei liegt der Fokus auf dem klassischen Zwillingsdesign, welches auch in KoSMoS realisiert wurde und auf dessen Grundlage die ätiologischen Analysen vorliegender Studie durchgeführt werden. Anschließend wird die verfügbare ätiologische Evidenz zu den hier relevanten Faktoren Schulerfolg, allgemeine kognitive Fähigkeit sowie Motivation zusammengefasst.

4.1 Grundlagen der Verhaltensgenetik

Innerhalb der Verhaltensgenetik werden zwei eigenständige Forschungstraditionen unterschieden: die molekulargenetische Forschung und die quantitativ verhaltensgenetische Forschung. Zu Beginn dieses Kapitels werden beide Ansätze in ihren Grundannahmen getrennt skizziert. Gemäß der Ausrichtung vorliegender Arbeit auf die Methode der quantitativen Verhaltensgenetik wird keine umfassende Betrachtung des molekulargenetischen Ansatzes vorgenommen. Zusammenfassende Darstellungen molekular-

genetischer Methodik und Forschungsbefunde finden sich beispielsweise bei Plomin, DeFries, McClearn und McGuffin (2008) sowie Riemann und F. M. Spinath (2005).

Molekulargenetik

Molekulargenetische Studien haben die Identifikation spezifischer Gene und Genloci zum Ziel, die eine Relevanz für interindividuelle Verhaltensunterschiede haben (F. M. Spinath & Deary, 2008). In diesem Forschungszweig der Verhaltensgenetik fand in den letzten Jahrzehnten vor allem eine sprunghafte methodische Entwicklung statt, wobei mittlerweile eine realistische Betrachtung ihrer Möglichkeiten und Grenzen Einzug gehalten hat (Nisbett et al., 2012; Riemann & F. M. Spinath, 2005).

Neben der Untersuchung monogenetischer Störungen finden sich primär Studien zu komplexen Merkmalen wie allgemeinen kognitiven Fähigkeiten, bei denen von einer polygenetischen Beeinflussung ausgegangen wird. Interindividuelle Unterschiede in diesen komplexen, kontinuierlich verteilten Merkmalen (z. B. Intelligenz) sind mit großer Wahrscheinlichkeit auf das Zusammenwirken zahlreicher Gene bzw. Genloci mit unterschiedlichen Effektgrößen zurückzuführen (Plomin et al., 2008; F. M. Spinath & Deary, 2008). Solche multiplen Gene, die zu einem quantitativen Merkmal beitragen, werden als *quantitative trait loci* (QTLs) bezeichnet. Durch Lokalisation dieser Genloci können Informationen darüber gewonnen werden, wie viele Gene an einem polygenetisch beeinflussten Merkmal mitwirken und ob sie in gleichem Ausmaß zu dem Merkmal beitragen oder ob einige Gene den Phänotyp stärker beeinflussen als andere (Klug, Cummings & Spencer, 2007). Die Identifikation von QTLs wird dabei durch die angenommenen geringen Effektstärken bei komplexen Merkmalen jedoch erheblich erschwert (Plomin et al., 2008).

In molekulargenetischen Untersuchungen werden in erster Linie zwei Ansätze verfolgt: *Linkage-Analysen* und *Allel¹-Assoziationsstudien*. Eine klassische Linkage-Analyse kann bei der Suche nach den Ursachen monogenetischer Störungen herangezogen werden. Dabei werden viele Familienstammbäume hinsichtlich der gemeinsamen Vererbung von Störung und spezifischem DNA-Marker über viele Generationen hinweg verfolgt (für eine genaue Beschreibung des Vorgehens siehe z. B. Riemann & F. M. Spinath, 2005). Um diesen Ansatz auch auf polygenetisch beeinflusste Merkmale anwenden zu können, wurde diese Technik grundlegend modifiziert (F. M. Spinath & Deary, 2008): Bei der Geschwisterpaar-Linkage-Analyse werden wenige Familienmitglieder (in der Regel Geschwister) über viele Familien hinweg untersucht. Allerdings wird auch bei dieser Methode angenommen, dass sie zur Aufdeckung von Effekten unter 10%, wie

¹ Allele: Einander entsprechende Erbanlagen homologer Chromosomen.

sie bei multifaktoriell beeinflussten Merkmalen zu erwarten sind, wenig geeignet ist (Plomin & Caspi, 1998).

Einen alternativen Zugang bietet die Allel-Assoziationsstudie, die typischerweise nicht auf Daten von verwandten Personen angewiesen ist, und die es ermöglicht, auch Geneinflüsse mit geringen Effektstärken zu identifizieren (Plomin et al., 2008). In diesem Ansatz wird eine Korrelation zwischen einem bestimmten Allel und einer Merkmalsausprägung in einer Populationsstichprobe berechnet, wobei neben einer Stichprobe von Merkmalsträgern auch eine Stichprobe von Kontrollpersonen erfasst wird (Riemann & F. M. Spinath, 2005).

Aktuelle Entwicklungen der Methoden in der molekulargenetischen Forschung wie z. B. Genpooling oder der Einsatz automatisierter Genotypisierung (z. B. mittels DNA-Chips) ermöglichen mittlerweile die gleichzeitige Berücksichtigung beider Vorgehensweisen (F. M. Spinath & Deary, 2008).

Quantitative Verhaltensgenetik

Die quantitative Verhaltensgenetik fragt nach den Ursachen von interindividuellen Unterschieden im Verhalten mit dem Ziel, die relative Bedeutung von genetischen Einflüssen und Umwelteinflüssen auf diese Unterschiede zu ermitteln.

Werden einzelne Merkmale untersucht, wird die beobachtete Variation in diesem Merkmal innerhalb einer Population in Quellen genetischer Variation und Umweltvariation zerlegt (Posthuma, 2009). Unter Verwendung genetisch informativer Designs (z. B. Adoptionsstudien oder den Vergleich gemeinsam aufgewachsener ein- und zweieiiger Zwillinge; vgl. Abschnitt 4.4), können die relativen Beiträge dieser genetischen und Umwelteffekte voneinander getrennt und deren Stärke quantifiziert werden. Bei Betrachtung mehrerer Merkmale erlauben genetisch informative Designs zusätzlich die Untersuchung der Ursachen der Kovariation zwischen diesen Merkmalen (siehe Kapitel 10.2). Solche multivariaten genetischen Analysen haben aufgrund ihrer größeren Anzahl an Interkorrelationen in der Regel eine größere Power als univariate genetische Analysen (Schmitz, Cherny & Fulker, 1998).

Nachfolgend werden die Grundlagen der quantitativen Verhaltensgenetik eingehend beschrieben. Zunächst wird das Konzept der Erbllichkeit und deren zugrunde liegenden Varianzquellen erläutert. Hieran anschließend erfolgt die Darstellung des in KoSMoS umgesetzten Designs gemeinsam aufgewachsener eineiiger (EZ) und zweieiiger Zwillinge (ZZ).

4.2 Abgrenzung von Anlage und Umwelt – Der Begriff der Erblichkeit

Der Begriff der *Erblichkeit* dient zur Beschreibung desjenigen Anteils der gesamten phänotypischen Variation in einer Population, der (im regressionsanalytischen Sinn) durch die Variation im Genotyp erklärt werden kann (Klug et al., 2007; Nisbett et al., 2012). Die formale Definition von Erblichkeit als Verhältnis der genetischen Varianz zur phänotypischen Varianz lautet daher (Borkenau, 1993):

$$H^2 = \frac{V_G}{V_P}$$

Es handelt sich hierbei also um das Verhältnis zweier Varianzen, wobei H^2 für die Erblichkeit, V_G allgemein für die genetische und V_P für die phänotypische oder beobachtete Varianz eines Merkmals steht. In der phänotypischen Varianz schlagen sich alle denkbaren Einflüsse (genetische und umweltbedingte) auf ein Merkmal nieder (Riemann & F. M. Spinath, 2005). Die Erblichkeit eines Merkmals gibt dementsprechend weder an, wie viel eines Merkmals genetisch bestimmt ist, noch sagt sie etwas über den Grad aus, zu dem der Phänotyp eines Individuums durch den Genotyp bestimmt wird (Klug et al., 2007). Erblichkeitskoeffizienten beziehen sich lediglich auf die jeweils untersuchte Population (die auch diversen Veränderungen und Stichprobenfehlern unterworfen sein kann) zu einer bestimmten Zeit in einer bestimmten Umwelt. Es handelt sich demnach nicht um Naturkonstanten, ein Umstand, der noch immer von Laien wie Wissenschaftlern häufig falsch verstanden und interpretiert wird (Borkenau, 1993; Klug et al., 2007; Riemann & F. M. Spinath, 2005). Hieraus folgt wiederum, dass es keinen Sinn macht, von *der* Erblichkeit eines Merkmals zu sprechen, denn diese hängt von der relativen Varianz der Prädiktoren (also Genotyp und Umwelt) ab und kann dementsprechend jeden Wert zwischen Null und Eins annehmen (Nisbett et al., 2012). Spezifischer wird nochmals zwischen *Erblichkeit im engen* und *Erblichkeit im weiten Sinne* unterschieden. Bei der Erblichkeit im engen Sinne geht lediglich die *additive genetische Varianz* in die betrachtete genetische Varianz ein, sie wird daher als h^2 bezeichnet (siehe Abschnitt 4.3). Sie wird im Forschungsdesign gemeinsam aufgewachsener EZ und ZZ geschätzt (siehe Abschnitt 4.4). Bei der Erblichkeit im weiten Sinne wird die gesamte genetische Varianz – bestehend sowohl aus additiven als auch epistatischen und dominanten Effekten (siehe Abschnitt 4.3) – ins Verhältnis zur phänotypischen Varianz gesetzt (H^2 ; Borkenau, 1993, Klug et al., 2007). Die Schätzung dieser Erblichkeit wird im Forschungsdesign getrennt aufgewachsener EZ vorgenommen. EZ teilen alle genetischen Effekte, sowohl die additiven als auch die nicht-additiven, vollständig (siehe Tabelle 4.1). Daher ist es beim Vergleich getrennt aufgewachsener EZ nicht möglich, additive Varianz und nicht-additive genetische Varianz zu trennen (Borkenau, 1993).

Borkenau (1993) weist außerdem darauf hin, dass ein hoher Erblchkeitskoeffizient eines Merkmals nicht mit dessen Unveränderbarkeit gleichzusetzen sei, dass aber eine hohe Erblchkeit mit gewissen Beschränkungen hinsichtlich der Veränderbarkeit dieses Merkmals durch Umweltfaktoren einhergeht.

4.3 Varianzzerlegung

Wie unter 4.1 bereits ausgeführt, setzt sich der Phänotyp jedes einzelnen Individuums aus genetischen Anteilen und umweltbedingten Anteilen zusammen. Als verhaltensbestimmend werden jedoch nicht nur genetische Effekte (G), sondern immer auch Effekte der Umwelt (E) auf den Phänotyp angenommen (Plomin et al., 2008). In diesem ist zudem der Messfehler enthalten. Algebraisch ergibt sich hieraus:

$$P = G + E$$

Konventionsgemäß wird mit P die Abweichung eines Individuums vom Populationsmittelwert bezeichnet (Plomin et al., 2008). In Anlehnung an Jensen (1969) und Riemann und F. M. Spinath (2005) ergibt sich folgende differenzierte Zerlegung der gesamten phänotypischen Varianz eines Merkmals:

$$V_P = \underbrace{[(V_{aG} + V_{AM}) + V_D + V_i]}_{\text{Erbanteil}} + \underbrace{[V_E + 2 \text{Cov}(H, E) + V_I]}_{\text{Umweltanteil}} + V_E$$

V_P	= phänotypische Varianz in der Population
V_{aG}	= genetische oder additive Varianz
V_{AM}	= Varianz aufgrund von selektiver Partnerwahl
V_D	= Varianz aufgrund von Dominanzabweichung
V_i	= Varianz aufgrund von Epistase
V_E	= Umweltvarianz (= Effekte der geteilten und nichtgeteilten Umwelt)
$\text{Cov}(H, E)$	= Kovarianz von Erbe und Umwelt
V_I	= Varianz aufgrund statistischer Interaktion zwischen genetischen und Umweltfaktoren
V_E	= Messfehler

Nachfolgend werden die einzelnen Varianzkomponenten im Detail beschrieben.

4.3.1 Genetische Varianz

Nach obiger Formel werden drei Arten genetischer Varianz unterschieden:

1. *Additive* genetische Varianz (V_{aG} oder a^2)
2. *Nicht-additive* genetische Varianz, die sich wiederum in Varianz aufgrund von Dominanzeffekten (V_D oder d^2) und Epistase (V_i oder i^2) aufgliedert und schließlich
3. Varianz aufgrund *selektiver Partnerwahl* (V_{AM} oder m^2)

Die *additive genetische Varianz* bezeichnet den Anteil der Merkmalsunterschiede, die durch die interindividuell unterschiedliche Wirkung von Genen an unterschiedlichen Loci erzeugt wird und ergibt sich somit als Summe der Wirkungen einzelner Gene (Amelang et al., 2006). Unter der Annahme der gleichen Wirkung aller an der Ausbildung eines Merkmals beteiligten Gene ergäbe sich im Idealfall eine perfekte lineare Beziehung zwischen der *Gendosis*² und der phänotypischen Merkmalsausprägung. Solche linearen Beziehungen sind nicht immer realistisch, trotzdem ist dieses Konzept innerhalb der Verhaltensgenetik sehr bedeutsam, denn die additiven Genwirkungen werden von Verwandten entsprechend einfacher Regeln geteilt. Dadurch lässt sich die additive genetische Varianz aus den Merkmalskorrelationen zwischen Verwandten eindeutig erschließen (Borkenau, 1993; vgl. auch Tabelle 4.1).

Würde die phänotypische Merkmalsvarianz ausschließlich V_{aG} enthalten, wäre der Mittelwert über alle Nachkommen eines Elternpaares gleich dem Durchschnittswert der phänotypischen Merkmalsvarianz der Eltern. Die mangelnde Übereinstimmung zwischen den Parental- und den Filialgenerationen kann durch Dominanzeffekte und Epistase erklärt werden (Borkenau, 1993). Als *Dominanzeffekte* werden solche Abweichungen zwischen dem erwarteten (additiven) und dem tatsächlichen genotypischen Wert bezeichnet, die auf die Interaktion der Allele am gleichen Genlocus zurückgehen, während *Epistase* auf interaktive Effekte zwischen Allelen an mehreren Genloci zurückgeht (Klug et al., 2007).

Schließlich entsteht Varianz auf Grund *selektiver Partnerwahl* (engl. *assortative mating*, AM), wenn in der Population keine zufällige Paarung der Partner erfolgt, sondern Korrelationen zwischen den Merkmalsausprägungen der Eltern vorliegen (Amelang et al., 2006).

² Die Gendosis bezeichnet die Anzahl merkmalspositiver Allele im Genotyp.

4.3.2 Umweltvarianz

Neben den genetischen Einflüssen werden auch Umwelteinflüsse (V_E) zur Erklärung interindividueller Differenzen in einem Merkmal herangezogen. Unter Umweltvarianz werden alle systematischen Effekte zusammengefasst, die nicht genetischer Herkunft sind, also biologische, soziale und kulturelle Faktoren (Amelang et al., 2006). Hier wird nochmals unterschieden zwischen geteilter (*common environment*, c^2) und nichtgeteilter Umwelt (*specific environment*, e^2). Effekte der *geteilten Umwelt* tragen zur Ähnlichkeit von gemeinsam aufwachsenden Individuen (z. B. den Mitgliedern einer Familie) bei. Als Beispiel wird häufig der SES der Familie genannt. Es handelt sich also um Umweltfaktoren, die von allen Mitgliedern einer Familie gleichermaßen geteilt werden und daher sowohl die Merkmalsvarianz als auch die Korrelation zwischen den Personen erhöhen.

Demgegenüber stehen Effekte der *nichtgeteilten Umwelt*, die ebenfalls zur Merkmalsvarianz, allerdings nicht zur Ähnlichkeit von Familienmitgliedern beitragen. Darunter fallen beispielsweise unterschiedliche Peergruppen oder Schullaufbahnen (Borkenau, 1993). Einen Überblick über die Verteilung der Varianzquellen in unterschiedlichen Verwandtschaftsbeziehungen liefert Tabelle 4.1.

Tabelle 4.1
Genetische Korrelationen und Umweltkorrelationen in verschiedenen Verwandtschaftsbeziehungen

Verwandtschaftsbeziehung	a^2	d^2	i^2	c^2
Eltern – Kind	0.50	0.00	0.00	1.00
Geschwister	0.50	0.25	0.00	1.00
Gemeinsam aufgewachsene EZ	1.00	1.00	1.00	1.00
Gemeinsam aufgewachsene ZZ	0.50	0.25	0.00	1.00
Getrennt aufgewachsene EZ	1.00	1.00	1.00	0.00
Getrennt aufgewachsene ZZ	0.50	0.25	0.00	0.00
Adoptiveltern–Adoptivkind	0.00	0.00	0.00	1.00
Adoptivgeschwister	0.00	0.00	0.00	1.00
Biologische Eltern–Adoptivkind	0.50	0.00	0.00	1.00

Anmerkungen. Adaptiert nach Riemann und F. M. Spinath, 2005, S. 562. a^2 = Effekte aufgrund additiver Genwirkungen; d^2 = Effekte aufgrund von Dominanzabweichung; i^2 = Effekte aufgrund von Epistase; c^2 = Effekte aufgrund geteilter Umwelt; EZ = Eineiige Zwillinge; ZZ = Zweieiige Zwillinge.

Bei der Abgrenzung von geteilten und nichtgeteilten Umwelteffekten kann es allerdings zu definitorischen Problemen kommen, denn jede erfasste Umweltvariable kann letztlich sowohl Effekte der geteilten als auch der nichtgeteilten Umwelt enthalten (Turkheimer

& Waldron, 2000). Es ist durchaus vorstellbar, dass eine Umweltvariable zwar als gemeinsamer Effekt angenommen wird – z. B. ein alleinerziehender Elternteil – er aber vollkommen unterschiedliche Auswirkungen auf die Kinder hat und damit eigentlich einen nichtgeteilten Effekt darstellt. Es ist daher sinnvoll, kindspezifische Auswirkungen zu untersuchen und sich nicht auf die Erfassung von Variablen auf dem Familienlevel zu beschränken (Turkheimer, D’Onofrio, Maes & Eaves, 2005).

Quantitativ verhaltensgenetische Studien haben in erster Linie das Wissen über Umwelteffekte erweitert bzw. das Verständnis von Umwelteffekten verändert. Zudem stammen die wichtigsten Belege für die Relevanz von Umwelteinflüssen zur Erklärung interindividueller Unterschiede aus der verhaltensgenetischen Forschung (Plomin, Asbury & Dunn, 2001; Plomin et al., 2008). Drei der wichtigsten Erkenntnisse verhaltensgenetischer Forschung beziehen sich denn auch auf die Umwelt und nicht auf die Genetik (vgl. Plomin et al., 2008):

1. Obwohl vielfach davon ausgegangen wird, dass sich die innerhalb der Familie bereitgestellte Umwelt auf die Entwicklung psychischer Merkmale von Kindern derart auswirkt, dass diese geteilten Erfahrungen zur Ähnlichkeit zwischen Geschwistern führt, konnte die verhaltensgenetische Forschung das Gegenteil belegen. Die Ergebnisse aus vielen Studien zeigen in konsistenter Weise, dass Umwelteinflüsse in erster Linie zur Unähnlichkeit und weniger zur Ähnlichkeit von Kindern, die in derselben Familie aufwachsen, beitragen (= Effekte der nichtgeteilten Umwelt; Plomin et al., 2001)
2. Weiterhin hat sich gezeigt, dass viele der in der psychologischen Forschung eingesetzten Umweltmaße selbst genetische Einflüsse aufweisen. Ergebnisse aus verschiedenen Studien legen den Schluss nahe, dass Menschen aktiv ihre eigenen Umwelten mitgestalten und dies teilweise aus genetischen Gründen (Plomin et al., 2008). Dieses Phänomen wird in der Verhaltensgenetik als *Gen-Umwelt-Korrelation* (r_{GE} , bzw. $\text{Cov}(H, E)$) bezeichnet, wurde aber weithin als *Nature of Nurture* (Genetik der Umwelt) bekannt.

Gen-Umwelt-Kovariation bzw. *-Korrelation* bezeichnet den Umstand, dass merkmalsrelevante genetische Faktoren und Umweltfaktoren nicht unabhängig voneinander sind. Genetisch determinierte Merkmale (z. B. kognitive Fähigkeiten) beeinflussen somit die Wahrscheinlichkeit, bestimmte Umwelterfahrungen zu machen, was wiederum auf das entsprechende Merkmal rückwirken könnte (Wolf & Riemann, 2008). Plomin, DeFries und Loehlin (1977) unterscheiden zwischen der passiven, der reaktiven und der aktiven Kovariation. Der *passive Typ* bezeichnet solche Arten der Kovariation, an deren Zustandekommen die Merkmalsträger nicht beteiligt sind. Als Beispiel hierfür nennt Borkenau (1993) den Umstand,

dass Kinder intelligenter Eltern eine größere Zahl merkmalspositiver Allele erben und ihnen ihre Eltern zugleich auch eine stärker Intelligenz fördernde Umwelt bieten. Die *reaktive Gen-Umwelt-Kovariation* drückt aus, dass Individuen nicht nur von ihrer Umwelt beeinflusst werden, sondern ihrerseits auch ihre Umwelt beeinflussen. So erlebt beispielsweise ein aggressives Kind in besonderem Maße ablehnende Reaktionen seiner Umwelt (Borkenau, 1993). Der *aktive Typ* umfasst schließlich diejenigen Fälle, in denen Personen ihre Umwelt durch ihr Verhalten auswählen und verändern können, z. B. durch bestimmte Hobbys oder die Berufswahl. Die Umwelt eines Individuums ist also zum Teil eine Konsequenz seines eigenen Verhaltens, welches wiederum durch den Genotyp beeinflusst sein mag.

3. Die dritte Erkenntnis in Bezug auf die Schnittstelle zwischen Anlage und Umwelt bezieht sich auf die sogenannte *Gen-Umwelt-Interaktion* (GxU , bzw. V_I). Diese beinhaltet die genetisch beeinflusste Empfänglichkeit für Umwelteinflüsse (Plomin et al., 2008). Es existieren verschiedene Vorstellungen über GxU (für einen Überblick siehe Rutter, 2006). Innerhalb der quantitativen Verhaltensgenetik bezeichnet sie jedoch den Umstand, dass Umwelteinflüsse in Abhängigkeit des Genotyps unterschiedliche Wirkungen entfalten (Asendorpf, 2004).

Zur Aufdeckung von GxU wurde z. B. auf Tier- oder Adoptionsstudien zurückgegriffen (für einen Überblick vgl. Plomin et al., 2008), aber auch die Zwillingsmethode wurde zur Identifikation von GxU herangezogen. Die verschiedenen Möglichkeiten der Modellierung einer GxU werden im Rahmen der vorliegenden Arbeit keine Rolle spielen, weshalb auf eine umfassende Darstellung an dieser Stelle verzichtet wird (eine Zusammenfassung des methodischen Vorgehens findet sich z. B. bei Spengler, 2009).

Zur Schätzung der zuvor beschriebenen Varianzquellen werden in der quantitativen Verhaltensgenetik Daten von Personen herangezogen, die in unterschiedlichen Verwandtschaftsverhältnissen zueinander stehen. Dabei wird sich die Tatsache zu Nutze gemacht, dass deren genetische und umweltbedingte Ähnlichkeit variiert (vgl. hierzu Tabelle 4.1) und deren phänotypisch beobachtete Ähnlichkeit zu deren genetischer Ähnlichkeit in Beziehung gesetzt werden kann. Der Hauptfokus liegt entsprechend auf der Kovarianz von Individuen in unterschiedlichem Verwandtschaftsverhältnis (Plomin et al., 2008). Hierzu werden Adoptions-, Familien- und Zwillingsdesigns, sowie die Kombination verschiedener Ansätze verwendet (Brocke et al., 2004; Plomin et al., 2008).

Im Folgenden wird das Forschungsdesign des Vergleichs gemeinsam aufgewachsener EZ und ZZ, welches auch als *klassisches Zwillingsdesign* (engl. *classical twin design*, CTD) bezeichnet wird, beschrieben. Für eine vertiefende Darstellung anderer Untersuchungsdesigns sei auf Borkenau (1993) sowie Plomin et al. (2008) verwiesen.

4.4 Das klassische Zwillingsdesign als Methode der quantitativen Verhaltensgenetik

Das CTD stellt das am häufigsten realisierte Design in der verhaltensgenetischen Forschung dar. Die Schätzung der Erbllichkeit bezieht sich hier auf die *Erblichkeit im engen Sinne* (vgl. Abschnitt 4.2).

Innerhalb des CTDs liegt die Annahme zugrunde, dass EZ (oder monozygote Zwillinge) exakt die gleichen Erbanlagen aufweisen, ihre genetischen Effekte (sowohl additive als auch nicht-additive) entsprechend zu 1.00 korrelieren. Demgegenüber beträgt die genetische Ähnlichkeit von ZZ (oder dizygote Zwillinge) bezogen auf solche Gene, die sich zwischen Menschen unterscheiden, durchschnittlich 50%. Sie sind sich somit genetisch genauso ähnlich wie normale Geschwister.

EZ teilen durch ihr identisches Erbgut sämtliche Effekte der Gendominanz und Epistasie, wodurch sich die Ähnlichkeit von EZ bei Vorliegen solcher Effekte erhöht. Bei ZZ korrelieren im Unterschied zu EZ Dominanzeffekte im Schnitt zu 0.25, epistatische Effekte in Abhängigkeit der beteiligten Loci in der Regel zu 0.00³. Beide nicht-additiven Effekte wirken sich demnach stärker auf die Ähnlichkeit von EZ-Paaren, als auf die Ähnlichkeit von ZZ-Paaren aus, was wiederum zu einer Überschätzung der Erbllichkeit im untersuchten Merkmal führen kann.

Für die Einflüsse der Umwelt wird davon ausgegangen, dass die geteilte Umwelt auf EZ und ZZ in gleicher Weise wirkt, deren Korrelation dementsprechend jeweils bei 1.00 liegt. Definitionsgemäß korrelieren die Effekte der nichtgeteilten Umwelt bei EZ und ZZ zu 0.00 (vgl. Abschnitt 4.3.2). Die Anwendung dieses Designs setzt eine möglichst exakte Bestimmung der Zygocie der Zwillinge voraus. Es gibt hierzu mehr oder weniger aufwendige und exakte Verfahren. Innerhalb der KoSMoS Studie wurde die Zygocie mittels Fragebogen ermittelt, die zu 96% eine exakte Bestimmung der Zygocie liefern (Price et al., 2000b; siehe Kapitel 8.1).

Voraussetzungen des CTD

Das Untersuchungsdesign gemeinsam aufgewachsener EZ und ZZ ist an verschiedene Vorannahmen geknüpft (Neale & Cardon, 1992).

Zunächst wird die Gültigkeit der sog. *Equal Environments Assumption* (EEA) angenommen. Diese besagt, dass die geteilte Umwelt EZ und ZZ gleichermaßen beeinflusst und sie somit zur Ähnlichkeit von EZ nicht stärker beiträgt als zur Ähnlichkeit von

³ Die Wahrscheinlichkeit, dass die Interaktion zwischen verschiedenen Genloci bei beiden Geschwistern in gleicher Weise auftritt, ist sehr gering, vgl. Plomin et al., 2008.

ZZ. Eine Verletzung dieser Annahme hätte einen systematischen Fehler in den Parameterschätzungen zur Folge: Die größere Ähnlichkeit der EZ im Vergleich zu ZZ könnte dann auch durch überdurchschnittlich ähnliche Umwelten bedingt sein, was insgesamt zu einer Überschätzung der genetischen Effekte führen würde. Die Gültigkeit der EEA wurde in verschiedenen Studien bestätigt (z. B. Borkenau, Riemann, Angleitner & Spinath, 2002; T. J. Bouchard & Propping, 1993; Klump, Holly, Iacono, McGue & Willson, 2000). EZ wachsen zwar durchaus in ähnlicheren Umwelten auf, allerdings stellt dies eher eine Folge und weniger eine Ursache ihrer Verhaltensähnlichkeit dar (Borkenau et al., 2002). Eine Verletzung der EEA läge zudem nur dann vor, wenn die größere Ähnlichkeit der Umwelt von EZ auch tatsächlich einen Einfluss auf das untersuchte Merkmal hätte. Es konnte jedoch gezeigt werden, dass eine ähnlichere Behandlung von EZ durch ihre Eltern nicht zu einer größeren Ähnlichkeit in den untersuchten Merkmalen führt und damit keine Verletzung der EEA vorliegt (Loehlin & Nichols, 1976; Morris-Yates, Andrews, Howie & Henderson, 1990).

In diesem Zusammenhang ist es auch bedeutsam, dass die Operationalisierung der Umwelt durch den Selbstbericht der Zwillinge als äquivalent für beide Gruppen angenommen werden kann (Mitchell et al., 2007). Die Faktormittelwerte liegen zwar für EZ in der Regel höher als für ZZ, die Kovarianzen und Faktorladungen zwischen beiden Gruppen unterscheiden sich jedoch nicht (Neale, Lubke, Aggen & Dolan, 2005). Unterschiedliche Mittelwerte in beiden Gruppen sprechen analog zu Loehlin und Nichols (1976) für die Wahrnehmung ähnlicherer Umwelten bei EZ, sie implizieren aber keine Invarianz der Messung der Umwelt bei EZ und ZZ (Neale et al., 2005; siehe auch Rijdsdijk & Sham, 2002 für eine ausführliche Diskussion der EEA).

Eine zweite Annahme innerhalb des CTD bezieht sich auf das Nichtvorliegen selektiver Partnerwahl, sowie von Kontrast- und Assimilationseffekten in der untersuchten Population.

Selektive Partnerwahl ist für die verhaltensgenetische Forschung aus zweierlei Gründen bedeutsam. Zum einen weichen die Nachkommen im Fall selektiver Partnerwahl stärker vom Mittelwert ab als bei zufälliger Paarung und vergrößern damit die Varianz dieses Merkmals in einer Population (Borkenau, 1993; F. M. Spinath et al., 2008a). Zum anderen führt die selektive Partnerwahl bei Verwendung der Zwillingsmethode zu einer Unterschätzung der Erbllichkeit (Plomin et al., 2008). Während selektive Partnerwahl auf Grund der genetischen Korrelation von 1.00 bei EZ keinen Einfluss auf deren Ähnlichkeit hat, führt sie bei ZZ zu einer Erhöhung der Ähnlichkeit der Geschwister, wodurch sich die Differenz zwischen EZ- und ZZ-Korrelationen verringert, die bei der Zwillingsmethode die Grundlage für Erbllichkeitsschätzungen darstellt (vgl. Abschnitt 4.5). Wird die selektive Partnerwahl nicht berücksichtigt, werden diese Effekte der geteilten Umwelt zugerechnet (Keller & Coventry, 2005). Im Hinblick auf Persönlichkeitsmerkmale wurden recht geringe Partnerkorrelationen zwischen 0.10 bis

0.20 gefunden (siehe Plomin et al., 2008). Die selektive Partnerwahl hinsichtlich g ist hingegen mit einer durchschnittlichen Höhe von 0.40 substantiell (Jensen, 1998). Liegen Daten der Eltern vor, kann eine direkte Schätzung des Effektes der selektiven Partnerwahl vorgenommen werden (siehe Keller et al., 2009).

Kontrast- und Assimilationseffekte können ebenfalls zu verzerrten Schätzungen von genetischen und Umwelteffekten führen. Beide Effekte sind vor allem bei Untersuchungen mit Selbst- oder Fremdbeurteilungen bedeutsam und können zu einer Überschätzung der genetischen Varianz führen. Kontrasteffekte entstehen vor allem dann, wenn Fremdbeurteilungen von Personen vorgenommen werden, die beide Zwillingsgeschwister kennen. Die Beurteiler verstärken dabei die Unterschiede zwischen den Geschwistern, weil sie diese vor allem untereinander und weniger mit anderen Personen in der Population vergleichen. Saudino, McGuire, Reiss, Hetherington und Plomin (1995) konnten nachweisen, dass der Kontrasteffekt für ZZ einflussreicher ist als für EZ, was eine Überschätzung des genetischen Einflusses zur Folge hat. In dieser Studie zeigte sich auch, dass bei EZ Assimilationseffekte eine bedeutendere Rolle spielen, d.h., dass Ähnlichkeiten zwischen EZ häufig besonders stark hervorgehoben werden. Eine Möglichkeit, um Problemen wie Kontrast- und Assimilationseffekten entgegenzuwirken, stellen Fremdbeurteilungen dar, bei denen die Beurteiler nur einen der beiden Zwillinge kennen.

Als weitere Voraussetzungen wird im CTD angenommen, dass keine Effekte der $G \times U$ und r_{GE} vorliegen, sondern sich die verschiedenen Einflüsse auf die Merkmalsvarianz additiv zusammensetzen und sich entsprechend ein linearer Effekt ergibt (Plomin et al., 2008). Beide Effekte werden im CTD nicht geschätzt, was gleichzeitig die umstrittenste Annahme des CTDs darstellt (siehe hierzu auch Kapitel 9.2)

Eine häufig angeführte Kritik an der Zwillingsmethode betrifft eine potenziell mangelnde Generalisierbarkeit der Ergebnisse auf die gesamte Population auf Grund einer höheren intrauterinen sowie familiären Vergleichbarkeit der Bedingungen für Zwillingsgeschwister. Diese Vermutung konnten Posthuma, De Geus, Bleichrodt und Boomsma (2000) für kognitive Fähigkeiten widerlegen.

Studien an gemeinsam aufgewachsenen Zwillingen stellen insgesamt ein wertvolles Instrument der verhaltensgenetischen Forschung dar. Verglichen mit Untersuchungen an getrennt aufgewachsenen EZ oder Adoptionsstudien können hier sehr viel größere Stichproben untersucht werden, wodurch die Repräsentativität der Stichprobe und die Generalisierbarkeit der Befunde auf die Population in stärkerem Maße gewährleistet sind. Als Hinweis auf die Güte von Untersuchungen an gemeinsam aufgewachsenen Zwillingen können auch Studien gewertet werden, welche die gleichen Konstrukte mit jeweils unterschiedlichem Design erfassen und ähnliche Schätzungen der Erblichkeit berichten

(vgl. z. B. T. Bouchard & McGue, 1981; Pedersen, Plomin, Nesselroade & McClearn, 1992).

Erweiterungen des klassischen Designs

Die rasche methodische Entwicklung innerhalb der quantitativen Verhaltensgenetik zeigt sich nicht zuletzt in der Möglichkeit, komplexere Modellierungen mit mehreren Verwandtschaftsgruppen vorzunehmen. Hierdurch kann die statistische Power der Analysen erhöht und damit auch eine größere Generalisierbarkeit der Ergebnisse erreicht werden (Björklund, Jäntti & Solon, 2005; Keller et al., 2009). Zudem wird es durch Hinzunahme weiterer Verwandtschaftsgruppen möglich, eine genauere Differenzierung der Varianzquellen vorzunehmen (simultane Modellierung additiver und nicht-additiver genetischer Effekte; siehe hierzu auch Kapitel 9.2; Keller et al., 2009).

Die Erweiterungen des CTDs können wie folgt beschrieben werden:

- Das *Nuclear Twin Family Design* (NTFD): Innerhalb dieses Designs werden die Eltern der Zwillinge in die Modellierung aufgenommen. Aufgrund der dadurch erreichten Erhöhung der Freiheitsgrade wird eine simultane Schätzung der Effekte der geteilten Umwelt sowie nicht-additiver genetischer Effekte möglich.
- Das *Extended Twin Family Design* (ETFD): Im ETFD wird die Gruppe der EZ und ZZ um Gruppen erweitert (z. B. Halbgeschwister oder Cousins), die in einem anderen Verwandtschaftsverhältnis zueinander stehen und entsprechend Tabelle 4.1, genetische und umweltbedingte Effekte teilen (Keller et al., 2009). Durch die Aufnahme von einer bzw. zwei zusätzlichen Gruppen erhöht sich die Power des Designs, was eine geringe Stichprobengröße zur Aufdeckung von Effekten notwendig macht (Posthuma et al., 2000).
- Das *Children of Twins Design* (COTD) erweitert die Modellierung um eine neue Generation, die Kinder der Zwillinge, wodurch der direkte Pfad von den Eltern zu den Kindern betrachtet werden kann (Eaves, Silberg & Maes, 2005).

Durch die zahlreichen neuen Möglichkeiten der simultanen Schätzung verschiedener Verwandtschaftsgruppen können in Abhängigkeit der untersuchten Fragestellung unterschiedliche Designs herangezogen werden. Die Wahl des Designs hängt demnach in erster Linie von der inhaltlichen Fragestellung, zusätzlich aber auch von der praktischen Umsetzbarkeit innerhalb der entsprechenden Studie ab.

4.5 Schätzung von genetischer Varianz und Umweltvarianz

Wie aus den bisherigen Ausführungen deutlich wurde, können verhaltensgenetische Studien genutzt werden, um sowohl die Bedeutung genetischer Varianzquellen, als auch die Wirkung der Umwelt auf die Ausprägung interindividueller Unterschiede aufzuzeigen. Die Stärke dieser Ansätze ist vor allem in der Möglichkeit zu sehen, genetische Effekte und Effekte der Umwelt in einem Forschungsdesign gemeinsam zu berücksichtigen, wodurch eine eindeutige Abgrenzung beider Effekte möglich wird.

Die folgende Darstellung bezieht sich auf das Design gemeinsam aufgewachsener EZ und ZZ. Trotz seiner relativ starken Vorannahmen (vgl. Abschnitt 4.4) bietet dieser Ansatz den entscheidenden Vorteil, dass Zwillinge die gesamte Bandbreite an Genotypen repräsentieren und auch hinsichtlich der Umwelten, in denen sie aufwachsen und leben, keine systematischen Einschränkungen zu erwarten sind (Riemann & F. M. Spinath, 2005).

Wie bereits erläutert beruht die Zwillingsmethode auf der Annahme, dass die Umweltvarianz bei EZ und ZZ gleich groß ist. Eine beobachtete größere phänotypische Ähnlichkeit der EZ kann demzufolge einzig auf deren größere genetische Ähnlichkeit zurückgehen. Auf der Grundlage von Ähnlichkeitskoeffizienten (Intraklassenkorrelation, ICC⁴) innerhalb der EZ Paare im Vergleich mit den Ähnlichkeitskoeffizienten innerhalb der ZZ Paare können bereits grobe Schätzungen der Erbllichkeit im engeren Sinne (h^2) und der Umwelteffekte vorgenommen werden. Erreichen EZ in einem Verhaltensmerkmal ungefähr doppelt so hohe ICCs wie ZZ und sind diese ICCs signifikant, spricht dies für eine hohe Erbllichkeit dieses Merkmals. Unterscheiden sich EZ und ZZ hingegen überhaupt nicht oder nur geringfügig in ihrer Ähnlichkeit, legt dies den Einfluss geteilter Umwelteffekte nahe. Bei niedrigen ICCs sowohl bei EZ als auch ZZ sind vorrangig nichtgeteilte Umwelteffekte als bedeutsam anzunehmen.

Die am häufigsten verwendeten Formeln zur Abschätzung der verschiedenen Varianzquellen gehen auf Falconer (1960) zurück. Im Fall des Vergleichs von gemeinsam aufgewachsenen EZ und ZZ kann die beobachtete Ähnlichkeit der EZ und ZZ Paare, den Annahmen des CTD folgend, in Form von geteilten Varianzkomponenten ausgedrückt werden als (Plomin et al., 2008):

$$ICC_{EZ} = h^2 + c^2$$

$$ICC_{ZZ} = \frac{1}{2}h^2 + c^2$$

⁴ Intraklassenkorrelationen spiegeln im Gegensatz zur Produkt-Moment-Korrelation auch Unterschiede bzgl. Streuung und Mittelwert wider (Riemann & F. M. Spinath, 2005).

Durch Subtraktion der ICC_{ZZ} von ICC_{EZ} ergibt sich als Schätzung der Erblichkeit:

$$h^2 = 2 (ICC_{EZ} - ICC_{ZZ})$$

Die Effekte der geteilten Umwelt lassen sich anhand der folgenden Formel bestimmen als:

$$c^2 = 2 ICC_{ZZ} - ICC_{EZ}$$

Zur Bestimmung der Effekte der nichtgeteilten Umwelt wird dem Umstand Rechnung getragen, dass dies die einzige Varianzquelle ist, die systematisch zur Unterschiedlichkeit von EZ beiträgt. Dementsprechend gilt:

$$e^2 = 1 - ICC_{EZ}$$

Die Schätzung der Effekte der nichtgeteilten Umwelt sind konfundiert mit dem Messfehler. Ist die Reliabilität der Messung bekannt, kann diese in der Formel berücksichtigt werden, um eine Erhöhung der Schätzgenauigkeit zu erreichen (Borkenau, 1993; Jang, McCrae, Angleitner, Riemann & Livesley, 1998).

Der Vergleich der Ähnlichkeit zwischen EZ und ZZ stellt einen ersten wichtigen Schritt beim Verständnis der Ätiologie verschiedener Merkmale dar. Allerdings trägt er Unterschieden in der Stichprobengröße von EZ und ZZ Paaren ebensowenig Rechnung, wie Unterschieden in Mittelwert und Varianz der Gruppen (F. M. Spinath & Johnson, 2011). In der Regel erfolgt die Analyse von Zwillingsdaten daher unter Rückgriff auf wesentlich flexiblere Strukturgleichungsmodelle und damit einhergehend auf die Bewertung des Modell-Fits der geschätzten Modelle (vgl. Kapitel 10). Strukturgleichungsmodelle stellen innerhalb sozialwissenschaftlicher Forschung einen verbreiteten Ansatz dar und bieten auch für die verhaltensgenetische Forschung gegenüber dem einfachen Vergleich der ICC zwischen EZ und ZZ mehrere Vorteile (Neale, 2009). Einerseits bieten sie die Möglichkeit, unterschiedliche Verwandtschaftsbeziehungen zur Schätzung von genetischen und Umwelteffekten simultan zu nutzen. Andererseits wird durch Fit-Statistiken die Güte der Übereinstimmung der empirischen Daten an die theoretische Modellvorstellung (im CTD der größeren Ähnlichkeit von EZ im Vergleich zu ZZ) angegeben und es können unterschiedliche Submodelle gegeneinander getestet werden (Neale & Cardon, 1992).

Allgemein besteht neben der univariaten Betrachtung der Ursachen interindividueller Unterschiede in bestimmten Merkmalen auch die Möglichkeit zur multivariaten Betrachtung von mehreren Merkmalen die zeitgleich, bzw. von Einzelmerkmalen die zu verschiedenen Zeitpunkten erfasst wurden. Hierdurch wird es möglich, die Ursachen der

Kovariation von Merkmalen aufzudecken (=bivariate Erblichkeit bzw. bivariate geteilte und nichtgeteilte Umwelt) und den überlappenden genetischen bzw. Umweltanteil zu bestimmen (=genetische Korrelation bzw. geteilte und nichtgeteilte Umweltkorrelation). Im Spezialfall einer längsschnittlichen genetischen Analyse eines Merkmals können zudem die Ursachen von Stabilität und Veränderung aufgedeckt werden.

Im Rahmen des CTD ist es allerdings nicht möglich, nicht-additive genetische Effekte und geteilte Umwelteffekte simultan in einem Modell zu testen. Eine ausführliche Darstellung des Vorgehens bei der Schätzung von Strukturgleichungsmodellen im Rahmen genetischer Analysen und eine Konkretisierung der in der vorliegenden Arbeit verwendeten Modelle wird in Kapitel 10 vorgenommen.

Ergänzend zur allgemeinen Darstellung der quantitativ verhaltensgenetischen Methoden werden im folgenden empirischen Teil solche Ergebnisse verhaltensgenetischer Forschung konkretisiert, die sich auf das hier untersuchte Kriterium Schulerfolg, sowie die Prädiktoren allgemeine kognitive Fähigkeiten und Motivation beziehen.

Abschließend werden solche Studien betrachtet, die beide Prädiktoren gemeinsam in Relation zu schulischem Erfolg untersuchen.

4.6 Befunde verhaltensgenetischer Studien

Im Vergleich zur Fülle phänotypischer Studien zur Vorhersage schulischen Erfolgs fällt die ätiologische Befundlage vor allem in Bezug auf das Kriterium Schulerfolg und den motivationalen Variablen insgesamt noch relativ spärlich aus, während sich für allgemeine kognitive Fähigkeiten bereits zahlreiche verhaltensgenetische Studien finden. Für die motivationalen Variablen FSK und IM beschränkt sich die vorliegende empirische Basis bisher auf die beiden Zwillingsstudien TEDS und KoSMoS.

4.6.1 Verhaltensgenetische Studien zu Schulerfolg

Geht es um die Ursachen interindividueller Unterschiede schulischen Erfolgs, kann zunächst festgestellt werden, dass sich nur wenige verhaltensgenetische Studien auf die ersten Schuljahre beziehen, obwohl diese eine wichtige Entwicklungsphase im Leben eines Kindes darstellen (vgl. Kapitel 1). Zum anderen liegt deren Fokus mehrheitlich auf standardisierten Leistungstest und Leistungen in für die Schule relevanten Bereichen, wie z. B. der Lesefähigkeiten oder mathematischen Fähigkeiten. Erst in jüngerer Zeit finden sich auch verhaltensgenetische Untersuchungen, welche die genetischen und umweltbedingten Einflüsse auf schulischen Erfolg operationalisiert durch Fachnoten

betrachten. Wie in Kapitel 1.1 dargestellt liefern jedoch gerade Lehrerurteile einen wichtigen Einblick in die schulische Leistung von Kinder, da hier auch wesentliche motivationale Merkmale und Persönlichkeitseigenschaften der Kinder beim Lernen und in der Interaktion mit anderen Kindern in die Beurteilung einfließen (Walker et al., 2004).

Aufgrund der relativ begrenzten Befundlage in Bezug auf fachspezifische Noten werden zunächst Studien betrachtet, die sich auf standardisierte Leistungstests und akademische Fähigkeiten beziehen. Auch an Hand dieser Ergebnisse lassen sich Rückschlüsse auf die zu erwartenden Ursachen interindividueller Differenzen in Schulnoten ziehen.

Was *Schulleistungstests* anbelangt findet sich über verschiedene Studien hinweg ein moderater Einfluss sowohl genetischer als auch geteilter Umwelteffekte in verschiedenen Altersbereichen (im Mittel bei ca. 50% bzw. 25%). Für Testleistungen im Lesen, Rechnen sowie sprachlichen Fähigkeiten liegen in einer Studie mit Kindern zwischen 6 und 12 Jahren ($N = 278$ Paare) sowohl der genetische Effekt als auch der Effekt der geteilten Umwelt im Mittel bei 40% (Thompson, Detterman & Plomin, 1991; siehe Kovas et al., 2007⁵). Ergebnisse einer niederländischen Studie, die sich auf den gleichen Altersbereich bezieht, sprechen für einen etwas höheren genetischen Effekt und einen niedrigeren Effekt der geteilten Umwelt (57% respektive 27%; Bartels, Rietveld, Van Baal & Boomsma, 2002b). Auch für das Highschool Alter werden vergleichbare genetische und geteilte Umwelteffekte für Testleistungen sowohl für Englisch als auch Mathematik berichtet ($a^2 = 0.40$; $c^2 = 0.30$; $N = 2164$ Paare; Loehlin & Nichols, 1976). Wainwright, Wright, Luciano, Geffen und Martin (2005) finden in einer Studie mit 15- bis 18-jährigen ($N = 390$ Paare) schließlich eine substanzielle Erblichkeit von 60% und einen lediglich geringen Anteil geteilter Umwelteffekte von ca. 10%. Aufgrund der breiten Operationalisierung von Fähigkeiten ist die Vergleichbarkeit dieser Studie mit den übrigen Ergebnissen allerdings eingeschränkt.

Im Bereich *spezifischer Fähigkeiten*, die für die schulische Leistung von Relevanz sind, liegen die umfangreichsten Erkenntnisse für die Lesefähigkeit vor. Aus verschiedenen Untersuchungen kann auch hier auf einen substanziellen genetischen Einfluss auf die Lesefähigkeit bzw. -unfähigkeit geschlossen werden (für einen Überblick vgl. Hart & Petrill, 2009). In einer Zusammenfassung mehrerer kleinerer Zwillingsstudien berichten beispielsweise Stromswold (2001) eine mittlere Erblichkeit von 40% und einen mittleren Effekt der geteilten Umwelt von 45%. Andere Studien erfassen die Leseleistung differenzierter im Hinblick auf unterschiedliche darin enthaltene Aspekte (z. B. buchstabieren oder phonologisches Verständnis). Hier bestätigt sich ebenfalls ein mittlerer

⁵ Die Angaben beziehen sich auf eine Reanalyse der Daten mit Alterskorrektur. Das Fehlen dieser Korrektur in der ursprünglichen Analyse von Thompson et al., 1991 führte zu einer Überschätzung des Anteils der geteilten Umwelt.

genetischer Effekt von 50%, wohingegen der geteilte Umwelteffekt in Abhängigkeit der betrachteten Fähigkeit erheblich schwankt (zwischen 17% und 40%) und damit insgesamt geringer ausfällt als bei globaler Erfassung der Lesefähigkeit (z. B. B. Byrne et al., 2005; Petrill, Deater-Deckard, Thompson, De Thorne & Schatschneider, 2006).

Im Bereich *mathematischer Fähigkeiten* liegen die Erblichkeitsschätzungen insgesamt in einem hohen Bereich (ca. 70%), während geteilte Umwelteffekte mit einem Anteil von 10% vernachlässigbar gering ausfallen (z. B. Alarcón, Knopik & DeFries, 2000; Light, Defries & Olson, 1998).

Zudem konnte in verschiedenen Studien nachgewiesen werden, dass der phänotypisch gefundene Zusammenhang zwischen kognitiven Fähigkeiten und Leistungen in standardisierten Tests größtenteils, d.h. zwischen 42% und 90%, durch genetische Einflüsse erklärt werden kann (z. B. Alarcón et al., 2000; Bartels et al., 2002b). Die berichteten genetischen Korrelationen zwischen kognitiven Fähigkeiten und unterschiedlichen Leistungstests variieren zwischen 0.40 und 0.85, während die geteilte Umweltkorrelation mit 0.00 bis 0.60 geringer ausfällt (Bartels et al., 2002b; Thompson et al., 1991). Generell scheint dabei die genetische Korrelation mit zunehmendem Alter höher und die geteilte Umweltkorrelation niedriger auszufallen.

Umfassende Ergebnisse in Bezug auf die Frage nach den Ursachen interindividueller Unterschiede in schulischen Leistungen liegen aus TEDS für die von Lehrern eingeschätzte akademische Leistung basierend auf den NC Kriterien vor. Wie aus Abbildung 4.1 hervorgeht, bleiben der genetische und der geteilte Umwelteffekt für *Englisch* und *Mathematik* stabil und für beide Fächer in etwa gleicher Höhe bestehen ($a^2 = 0.63 - 0.72$; $c^2 = 0.04 - 0.17$). Für *Naturwissenschaften*, das lediglich für die Altersstufen 9 und 10 vorliegt, liegt der genetische Anteil im Alter von 10 Jahren zwar niedriger als mit 9 Jahren, der Unterschied ist allerdings nicht signifikant ($a^2 = 0.48$ bzw. 0.63). In Anbetracht der großen Veränderungen in der Schulumwelt und in den Bewertungsrichtlinien des NCs, gerade im Alter zwischen 7 und 10 Jahren, war diese hohe Stabilität in den Erblichkeitsschätzungen nicht unbedingt zu erwarten (Kovas et al., 2007). Um diesen Umstand näher zu beleuchten wird von Kovas et al. (2007) ein Vergleich mit ebenfalls aus TEDS vorliegenden Daten zur Erblichkeit kognitiver Fähigkeiten, bei denen ein Anstieg der Erblichkeit über die Zeit als gesichert gelten kann (siehe hierzu 4.6.2), herangezogen. Im entsprechenden Zeitintervall von drei Jahren finden sich, trotz unterschiedlicher Erfassungsmethode (Telefon, Testhefte und Online Testung; siehe Davis, Arden & Plomin, 2007), auch für kognitive Fähigkeiten keine quantitativen Veränderungen in der Höhe der Parameterschätzungen (0.36 mit 7 Jahren, 0.36 mit 9 Jahren und 0.41 mit 10 Jahren). Dies legt den Schluss nahe, dass die Stichprobe über eine zu

geringe Power verfügt, um Alterseffekte in diesem kurzen Zeitintervall adäquat abbilden zu können (Kovas et al., 2007).⁶

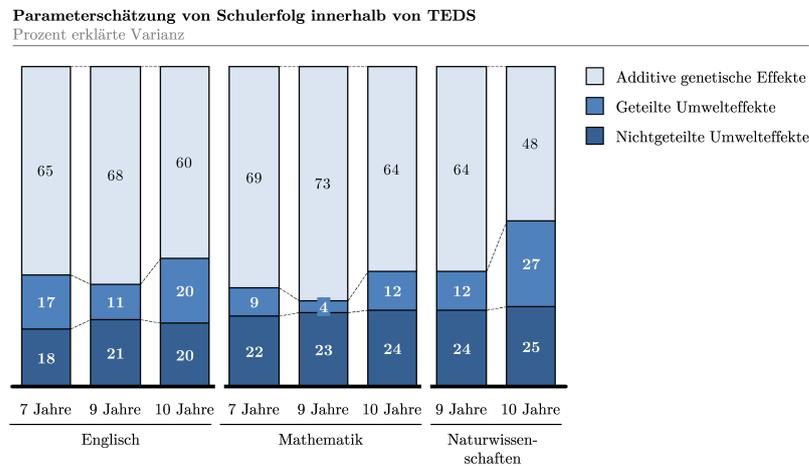


Abbildung 4.1. Erblichkeit schulischen Erfolgs in verschiedenen Altersbereichen. Eigene Darstellung auf Basis der in Kovas et al., 2007 berichteten Ergebnisse.

Neben dieser querschnittlichen Betrachtung quantitativer Unterschiede in verschiedenen Altersbereichen wurde innerhalb von TEDS in einer längsschnittlichen Betrachtung zudem eine Analyse der Ursachen von Stabilität und Veränderung schulischer Leistung vorgenommen. Sowohl für *Englisch* als auch *Mathematik* findet sich ein starker genereller genetischer Faktor. Dies bedeutet, dass jeweils ca. 50% der genetischen Gesamtvarianz im Alter von 9 und 10 Jahren mit der genetischen Varianz im Alter von 7 Jahren überlappt, was für eine hohe genetische Kontinuität spricht (Kovas et al., 2007). Der Großteil an verbleibender genetischer Varianz ist altersspezifisch, genetische Effekte können dementsprechend auch als Erklärung von Veränderung angenommen werden. Ca. 12% der spezifischen genetischen Varianz im Alter 9 trägt zudem zur Kontinuität im Alter 10 bei. Für das Schulfach *Naturwissenschaften* sprechen die Ergebnisse eher für eine genetische Veränderung denn für Kontinuität (Kovas et al., 2007).

Der Anteil an geteilter Umwelt liegt generell in einem niedrigen Bereich (über alle Altersstufen und Schulfächer hinweg bei 0.14), aber auch die geteilte Umwelt trägt sowohl zur Kontinuität als auch zur Veränderung bei (Kovas et al., 2007). Im Gegensatz dazu ist der Anteil der nichtgeteilten Umwelt in sämtlichen Schulfächern vollständig altersspezifisch und trägt damit lediglich zur Veränderung und nicht zur Stabilität bei. Allerdings ist hierin auch der Messfehler enthalten, was als Veränderung fehlinterpretiert werden könnte (Kovas et al., 2007). Abhilfe könnte hier die Verwendung eines latenten genetischen Modells, in dem der Messfehler und der nichtgeteilte Umweltan-

⁶ Die Differenz zwischen den Erblichkeitsschätzungen müsste in dieser Stichprobe 0.20 übersteigen, um signifikante Unterschiede aufzuzeigen (Kovas et al., 2007).

teil getrennt modelliert werden, schaffen.

Eine weitere Frage, die innerhalb von TEDS untersucht wurde, bezieht sich auf die Höhe der genetischen Korrelation, sowie der Umweltkorrelation zwischen den verschiedenen Altersstufen in den jeweiligen Fächern. Sowohl die genetischen Korrelationen als auch die geteilten Umweltkorrelationen liegen in einer substantiellen Höhe (zwischen 0.52 und 0.71), während die nichtgeteilten Umweltkorrelationen generell etwas niedriger ausfallen (Kovas et al., 2007). Schließlich wurde überprüft, auf welche Effekte die phänotypisch gefundenen Korrelationen zurückzuführen sind. Es zeigte sich für alle Fächer, dass die phänotypische Korrelation zwischen den drei Altersbereichen in erster Linie auf genetische Effekte zurückzuführen ist (zwischen 54% und 79%). Die Substichprobe der 9-jährigen wurde von F. M. Spinath et al. (2008a) im Hinblick auf potenzielle Geschlechtsunterschiede untersucht, es konnten jedoch weder quantitative noch qualitative Unterschiede bestätigt werden.

Einschränkend sei an dieser Stelle darauf hingewiesen, dass sich die zuvor berichteten hohen Erblichkeitsschätzungen auch teilweise darauf zurückführen lassen, dass die meisten Zwillingspaare vom gleichen Lehrer eingeschätzt wurden. Walker et al. (2004) führten eine Analyse auf Grundlage einer Substichprobe aus TEDS von 399 7-jährigen Zwillingspaaren durch, die von zwei verschiedenen Lehrern in den NC Kriterien eingeschätzt wurden. Es bestätigt sich zwar ein substantieller genetischer Effekt, dieser liegt jedoch mit im Schnitt $a^2 = 0.42$ in einem niedrigeren Bereich als bei Kovas et al. (2007) berichtet. Die Autoren führen dies auf unterschiedliche Ergebnisse für gleich- und gegengeschlechtliche ZZs zurück: Schätzen die gleichen Lehrer ein, fallen die ICCs für gleichgeschlechtliche Paare weitaus höher aus als für gegengeschlechtliche Paare (0.65 vs. 0.48). Diese Differenzierung findet sich bei der Einschätzung durch verschiedene Lehrer nicht (0.50 vs. 0.54).

Studien zur Erblichkeit von *Schulnoten* belegen im Vergleich zu standardisierten Schulleistungstests eine höhere Erblichkeit und einen deutlich geringeren Effekt der geteilten Umwelt (im Mittel ca. 60% bzw. 20%), was evt. dadurch zu erklären ist, dass Lehrer auch phänotypische Merkmale erfassen, die bei Leistungstests nicht berücksichtigt werden (Walker et al., 2004). Im Alter von 13 Jahren findet sich beispielsweise für Zeugnisnoten in den Fächern Lesen, Schreiben und Rechnen eine mittlere Erblichkeit von 50% und ein Effekt der geteilten Umwelt von 30% (Husén, 1959). Aktuellere Ergebnisse liegen aus der *Minnesota Twin Family Study* (MTFS; Iacono & McGue, 2002) vor. In der Kohorte der 11-jährigen (EZ=513; ZZ=287) liegt die mittlere Erblichkeit der Schulnote in verschiedenen Fächern bei 75% und damit weitaus höher als die berichteten Erblichkeiten für das in TEDS eingesetzte NC Maß (Johnson, McGue & Iacono, 2005). Der Effekt der geteilten Umwelt liegt im Mittel bei 19%, wobei sich keine Unterschiede in der Höhe der Parameterschätzungen für Jungen und Mädchen ergaben (Johnson et al., 2005). Die gleiche Stichprobe wurde auch längsschnittlich in

einem Follow-up mit 14 und 17 Jahren betrachtet. Der mittlere genetische Effekt für Mädchen liegt hier bei 0.66, für Jungen etwas niedriger bei 0.58. Auch die beobachtete lineare Veränderung der Schulnoten über die Zeit unterliegt einem starken genetischen Einfluss (70%; Johnson, McGue & Iacono, 2006).

Ergebnisse der KoSMoS Studie (N=280 Paare, durchschnittliches Alter: 13.1, SD=0.87) belegen für Schulnoten in Mathematik und Deutsch moderate genetische Effekte (0.47 bzw. 0.31), sowie Effekte der geteilten Umwelt in vergleichbarer Höhe (0.47 bzw. 0.31; Gottschling, Spengler, Spinath & Spinath, 2012).

Ein mittlerer genetischer Effekt auf Schulnoten (0.52) findet schließlich auch in der *Sibling Interaction and Behavior Study* (Johnson et al., 2007) Bestätigung. Der Effekt der geteilten Umwelt fällt in dieser Studie mit 0.17 deutlich geringer aus. Außerdem konnte eine signifikante Überlappung (51%) genetischer Einflüsse auf Intelligenz und Schulleistung aufgezeigt werden (Johnson et al., 2007).

4.6.2 Verhaltensgenetische Studien zu allgemeinen kognitiven Fähigkeiten

Wohl kaum ein anderes psychologisches Konstrukt wurde in der Geschichte der quantitativen Verhaltensgenetik so intensiv untersucht wie *kognitive Fähigkeiten*, und auch im Bereich molekulargenetischer Forschung liegen mittlerweile zahlreiche Studien vor. Der Fokus der nachfolgenden zusammenfassenden Darstellung liegt auf Erkenntnissen, die im Rahmen quantitativ verhaltensgenetischer Studien gewonnen wurden. Für eine Beschreibung der Befunde molekulargenetischer Studien sei auf Posthuma und De Geus (2006) sowie Plomin et al. (2008) verwiesen. Am Ende dieses Abschnittes werden schließlich Ergebnisse aus Studien dargestellt, die auch Gen-Umwelt-Interaktionen berücksichtigen.

Quantitativ verhaltensgenetische Studien mit insgesamt mehr als 10.000 Zwillingspaaren belegen einheitlich eine substanzielle genetische Beeinflussung allgemeiner kognitiver Fähigkeiten in mittlerer Höhe ($h^2 \approx 0.50$; vgl. z.B. Überblicksarbeiten von T. Bouchard & McGue, 1981; Deary, Spinath & Bates, 2006; Plomin & F. M. Spinath, 2004; Segal & Johnson, 2009; F. M. Spinath & Deary, 2008). Auch eine Studie, in der simultan Daten aus Familien-, Adoptions- und Zwillingsstudien modelliert wurden, bestätigt, dass interindividuelle Unterschiede in kognitiven Fähigkeiten zu ca. 50% durch genetische Effekte erklärt werden (Petrill et al., 2004). Vermutungen, dass es sich bei einer durchschnittlichen Erblichkeitsschätzung in dieser Höhe womöglich eher um eine Unterschätzung handelt, gehen zum einen darauf zurück, dass die Schätzungen der Erblichkeit höher ausfallen, wenn um die Unreliabilität der Messung korrigiert wird (Plomin et al., 2008). Zum anderen ist auch die selektive Partnerwahl zu berücksichtigen, denn die Korrelation zwischen Ehepartnern im Merkmal Intelligenz liegt mit

Werten zwischen 0.30 und 0.40 vergleichsweise hoch (T. Bouchard & McGue, 1981; Plomin et al., 2008). Dies kann zu einer Überschätzung der Ähnlichkeit vom ZZ und damit zu einer Überschätzung der Effekte der geteilten Umwelt führen (vgl. Abschnitt 4.3). Werden in Strukturgleichungsanalysen die Modellannahmen (z. B. durch Erhöhung der genetischen Ähnlichkeit zweieiiger Zwillinge auf > 0.50) entsprechend angepasst, steigen die Erblichkeitsschätzungen an (F. M. Spinath et al., 2008a). In einer aktuellen Studie, in der sowohl Effekte selektiver Partnerwahl als auch der kulturellen Transmission in einem *ETFD* berücksichtigt wurden, werden zudem neben additiven genetischen Effekten (44%) und nichtgeteilten Umwelteffekten (18%), nicht-additive genetische Effekte (27%) und Effekte selektiver Partnerwahl (11%) als erklärende Varianzquellen kognitiver Fähigkeiten identifiziert (Vinkhuyzen, van der Sluis, Maes & Posthuma, 2012). Dieses Befundmuster bedarf allerdings weiterer Studien, um deren empirische Evidenz zu stützen.

In den letzten Jahren finden sich darüber hinaus zunehmend Forschungsarbeiten, die sich auf entwicklungsbezogene Fragestellungen, beispielsweise auf die Frage, ob sich die Höhe der Erblichkeit im Laufe der Entwicklung verändert, konzentrieren (F. M. Spinath & Deary, 2008). Aus den Ergebnissen dieser Studien kann geschlussfolgert werden, dass eine alleinige Betrachtung der mittleren Erblichkeit zu kurz greift, um die Frage nach den Ursachen interindividueller Unterschiede in kognitiven Fähigkeiten angemessen zu beantworten, denn hierin spiegelt sich auch eine Populationsabhängigkeit der Erblichkeitsschätzungen wider (vgl. Abschnitt 4.2).

Die Ergebnisse zahlreicher (Querschnitt-)Studien sind in Abbildung 4.2 im Überblick dargestellt. Wie aus dieser Zusammenfassung hervorgeht, nimmt die Erblichkeit kognitiver Fähigkeiten im Laufe der Entwicklung zu, während der Einfluss der geteilten Umwelt kontinuierlich sinkt (Bartels et al., 2002b; Bergen, Gardner & Kendler, 2007; McGue, Bouchard, Iacono & Lykken, 1993). Liegt die Erblichkeit in der frühen Kindheit (2 bis 4 Jahre) noch in einem niedrigen Bereich bei ca. 30% (F. M. Spinath et al., 2003), steigt diese auf über 50% im frühen Erwachsenenalter (z. B. Finkel, Pedersen, McClearn, Plomin & Berg, 1996; Finkel & McGue, 1998; Haworth et al., 2010; Pedersen et al., 1992; Posthuma, De Geus & Boomsma, 2001) und liegt im mittleren und hohen Erwachsenenalter bei ca. 70-80% (z. B. McClearn et al., 1997; McGue & Christensen, 2001; Reynolds et al., 2005). Parallel dazu verliert die geteilte Umwelt im Verlauf der Entwicklung immer mehr an Bedeutung, im Erwachsenenalter werden diese Effekte nicht mehr signifikant (Bergen et al., 2007; Posthuma et al., 2001; Rijdsdijk, Vernon & Boomsma, 2002). Über die Lebensspanne hinweg gesehen sind Umwelteinflüsse auf kognitive Fähigkeiten dementsprechend hauptsächlich nichtgeteilter Art (Plomin & F. M. Spinath, 2004).

Haworth et al. (2010) fassen beispielsweise in einer Analyse Daten aus sechs Zwillings-

Parameterschätzung der Intelligenz je Altersgruppe
 Prozent erklärte Varianz

Gesamt-N: ca. 17.500 Zwillingspaare

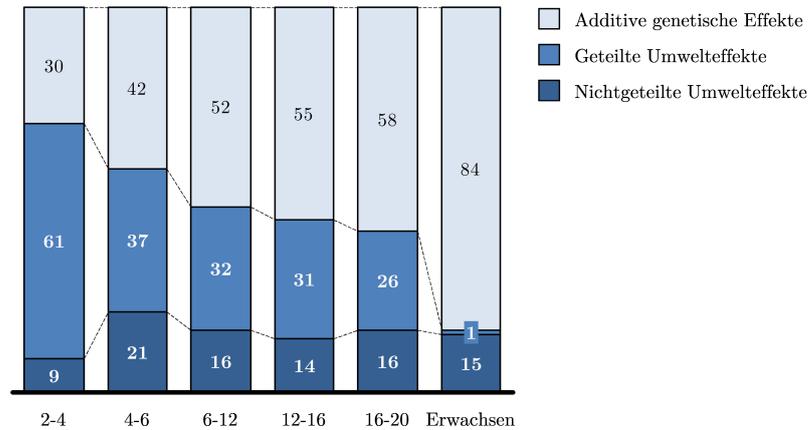


Abbildung 4.2. Erblichkeit der Intelligenz über die Lebensspanne hinweg. Eigene Darstellung in Anlehnung an Abbildung 2 aus Plomin und F. M. Spinath, 2004, S. 115, ergänzt um Daten aus F. M. Spinath et al., 2003.

studien aus vier unterschiedlichen Ländern (USA, UK, Niederlande und Australien) mit insgesamt 11.000 Zwillingspaaren im Alter zwischen 6 und 71 Jahren⁷ zusammen. In den drei gebildeten Altersgruppen *Kindheit* (4-10 Jahre; MW=9 Jahre), *Jugend* (11-13 Jahre, MW=12 Jahre) und *junges Erwachsenenalter* (14-34 Jahre; MW=17 Jahre) steigt die gefundene Erblichkeit des generierten *g-Faktors* linear von 41% auf 55% und schließlich auf 66% an. Weiterhin bleibt der Anteil der nichtgeteilten Umwelt mit ca. 25% bis zur Jugend stabil, sinkt dann jedoch auf 19% im jungen Erwachsenenalter. Ein entgegengesetzter Verlauf zeichnet sich bei Effekten der geteilten Umwelt ab: Diese sinken von der Kindheit bis zur Jugend relativ stark (von 33% auf 18%) und bleiben dann auf diesem mittleren Level stabil bis ins junge Erwachsenenalter.

Zusätzliche Bestätigung der wachsenden Bedeutung genetischer Einflüsse auf Intelligenz mit steigendem Alter liefert eine längsschnittliche Adoptionsstudie, die einen stetigen Anstieg genetischer Einflüsse auf *g* zwischen dem 3. und 16. Lebensjahr findet (Plomin, Fulker, Corley & DeFries, 1997).

Gleichwohl liegen auch Hinweise auf eine wieder sinkende Erblichkeit im hohen Erwachsenenalter vor. Basierend auf Daten aus dem schwedischen Zwillingregister (N=110 EZ und 130 gleichgeschlechtliche ZZ; durchschnittliches Alter=82) werden genetische Effekte von 0.62 für einen generellen Intelligenzfaktor und Verarbeitungsgeschwindigkeit, 0.55 für verbale Fähigkeiten, 0.32 für räumliche Fähigkeiten und 0.52 für Gedächtnis berichtet (McClearn et al., 1997). Dieses Absinken der Erblichkeit legt den Schluss nahe, dass Umweltfaktoren, insbesondere die physische Gesundheit, am Ende der Lebenszeit nochmal an Wichtigkeit gewinnen (Segal & Johnson, 2009).

⁷ In die endgültige Analyse gingen nur Daten von Zwillingen im Alter zwischen 6 und 34 Jahren ein, da die Stichprobengröße für ältere Paare mit n=127 zu gering ausfiel.

Für die Variation genetischer Einflüsse über die Lebensspanne hinweg werden unterschiedliche Ursachen diskutiert, deren Effekte sich auch gegenseitig überlagern könnten (Bergen et al., 2007). Der Anstieg der Erbllichkeit könnte zum einen dadurch erklärt werden, dass im Verlauf des Lebens zusätzliche Gene verhaltenswirksam werden und kognitive Fähigkeiten beeinflussen (Bergen et al., 2007; Plomin & F. M. Spinath, 2004). Diese Erklärung wird vor allem vor dem Hintergrund einer mit dem Alter ansteigenden Komplexität kognitiver Anforderungen (Riemann & F. M. Spinath, 2005) und dem Umstand plausibel, dass sich wesentliche strukturelle Veränderungen des Gehirns von der frühen zur mittleren Kindheit und während der Jugend vollziehen (Haworth et al., 2010).

Eine weitere Argumentation fokussiert auf die Bedeutung der aktiven r_{GE} , wonach Individuen im Lebensverlauf zunehmend Kontrolle über ihre Umwelt gewinnen und immer aktiver in die Auswahl und Ausgestaltung ihrer Umwelt eingreifen (Haworth et al., 2010; Scarr & McCartney, 1983). Durch die ihrerseits teilweise genetische Beeinflussung der Selbstselektion von Umwelten, Situationen und damit letztlich auch Erfahrungen sollten die Erbllichkeitsschätzungen für kognitive Fähigkeiten über die Lebensspanne hinweg ansteigen (Riemann & F. M. Spinath, 2005). Auf den Vergleich von EZ und ZZ übertragen würde aufgrund deren größerer genetischer Ähnlichkeit eine aktive r_{GE} bei EZ mit größerer Wahrscheinlichkeit zur Wahl gleichartiger Umwelten führen als bei ZZ (Bergen et al., 2007). Dies würde sich für EZ in konsistent hohen bzw. steigenden ICCs widerspiegeln, während sich ZZ immer unähnlicher werden sollten. In der Tat findet sich dieser Trend bei Betrachtung der Zwillingsähnlichkeiten über die Lebensspanne: Während sich EZ in Intelligenzmaßen schon von frühester Kindheit an sehr ähnlich sind, kann bei ZZ ab dem jungen Erwachsenenalter ein deutliches Absinken der Ähnlichkeit beobachtet werden (vgl. Plomin & F. M. Spinath, 2004, S.115, Abbildung 2).

Schließlich kann auch eine Erklärung in Betracht gezogen werden, die auf die Berechnung genetischer und Umweltanteile abzielt. Verringert sich generell die Varianz der Umwelteinflüsse, steigt zwangsläufig die Schätzung der Erbllichkeit an, denn die beobachteten phänotypischen Unterschiede können im Fall gleicher Umwelten nur auf genetischen Unterschieden beruhen. Dies wäre einerseits dann der Fall, wenn geteilte Umwelteffekte oder die passive r_{GE} an Bedeutung verlören oder andererseits, wenn sich die Fehlervarianz der Messung verringerte (z. B. durch eine steigende Fähigkeit, sein Verhalten zu reflektieren und selbst zu beurteilen; Bergen et al., 2007).

Im Rahmen *multivariater genetischer Analysen* wurde auch die genetische Korrelation⁸ spezifischer kognitiver Fähigkeiten (z. B. verbale und räumliche Fähigkeiten) un-

⁸ Diese bezeichnet das Ausmaß, in dem genetisch bedingte Prozesse, die auf ein Merkmal wirken,

tersucht. Univariat betrachtet fallen die genetischen Einflüsse auf diese Fähigkeiten in der Regel zwar etwas geringer aus als für allgemeine Intelligenz (vgl. z. B. Kovas et al., 2007; Plomin & DeFries, 1998), sind jedoch ebenfalls substanziell. Die genetische Korrelation zwischen verschiedenen kognitiven Fähigkeiten liegt üblicherweise in einem sehr hohen Bereich und erreicht Werte bis zu 1.00 (Pedersen & McClearn, 1994; Petrill, Saudino, Wilkerson & Plomin, 2001; Price et al., 2000a), was von den Autoren dahingehend interpretiert wird, dass die gleichen Gene verschiedene kognitive Fähigkeiten determinieren (eine kritische Betrachtung dieser Interpretation der genetischen Korrelation wird von Johnson, Penke & Spinath, 2011 gegeben; siehe dazu auch Kapitel 10.2). Darüber hinaus fanden sich hohe bivariate Erblichkeiten zwischen verbalen und nonverbalen Testaufgaben und eine moderate bivariate geteilte Umwelt (Reznick, Corley, Robinson & Matheny, 1997). Phänotypisch gefundene Korrelationen sind dementsprechend v.a. durch genetische Effekte vermittelt. Selbst der phänotypisch gefundene Zusammenhang zwischen psychometrischer Intelligenz und der Informationsverarbeitungsgeschwindigkeit ist weitgehend genetisch vermittelt (Neubauer, Spinath, Riemann, Borkenau & Angleitner, 2000). Auf Grundlage dieser Befunde wird analog zu einem psychometrischen oder phänotypischen *g*-Faktor auch von einem *genetischen g-Faktor* gesprochen, der die genetische Varianz fast vollständig erklärt (Plomin & F. M. Spinath, 2002; Rijdsdijk & Sham, 2002). Dieser genetische *g*-Faktor konnte von D. Luo et al. (1994) bereits in der mittleren Kindheit nachgewiesen werden.

Ein weiterer Aspekt multivariater Betrachtung betrifft die Frage nach den Ursachen von Stabilität und Veränderung in allgemeinen kognitiven Fähigkeiten (zur phänotypischen Stabilität siehe Kapitel 2.4). Diesbezüglich konnte in verschiedenen Studien gezeigt werden, dass die Mehrzahl genetischer Effekte auf allgemeine kognitive Fähigkeiten zu deren Stabilität beitragen (z. B. Bartels, Rietveld, Van Baal & Boomsma, 2002a; Bishop et al., 2003; Petrill et al., 2004; Rietveld, Dolan, vanBaal & Boomsma, 2003). In geringerem Ausmaß finden sich für jede Altersstufe auch neue genetische Effekte, was dafür spricht, dass genetische Effekte zusätzlich einen Beitrag zur Veränderung in diesem Merkmal leisten (Plomin et al., 2008). Für nichtgeteilte Umwelteffekte weisen Ergebnisse aus Längsschnittanalysen hingegen darauf hin, dass diese als altersspezifische Effekte auftreten und damit nur zur Veränderung beitragen (Bartels et al., 2002a; Bishop et al., 2003).

Schließlich wurde auch der Zusammenhang zwischen allgemeinen kognitiven Fähigkeiten und Schulerfolg in multivariaten genetischen Analysen betrachtet. Für TEDS berichten Kovas et al. (2007) separate Ergebnisse für die Altersgruppen 7, 9 und 10 Jahre. Für alle Altersbereiche zeigte sich konsistent, dass die genetischen Effekte auf akademische Leistungen nicht vollständig deckungsgleich mit den genetischen Effekten

mit den genetischen Prozessen, die auf ein anderes Merkmal wirken, überlappen (Plomin et al., 2008). Siehe dazu Kapitel 10.

auf den *g-Faktor* sind. Für den *NC* in Englisch und Mathematik und den *g-Faktor* fällt die genetische Korrelation mit ca. 0.60 substantiell aus. Ein differierendes Bild zeigt sich für Effekte der geteilten Umwelt, die für *g* und akademische Leistungen nur wenig überlappen (diese überlappen zwischen verschiedenen Schulfächern allerdings zu einem großen Teil). Die Effekte der nichtgeteilten Umwelt sind hingegen spezifisch sowohl für den *g-Faktor* als auch die Leistung in jedem Schulfach. Die phänotypischen Korrelationen zwischen *g* und den *NC* Bewertungen sind hauptsächlich genetisch vermittelt ($biv_{h^2} = 76\%$; $biv_{c^2} = 18\%$; $biv_{e^2} = 6\%$). In der KoSMoS Stichprobe fällt die genetische Korrelation für 13-jährige zwischen Schulnoten in Mathematik und Deutsch und dem gebildeten *g-Faktor* mit 0.81 respektive 0.89 noch wesentlich höher aus, während die Umweltkorrelationen keine Signifikanz erreichen (Gottschling et al., 2012).

Abschließend werden Ergebnisse aus aktuellen Gen-Umwelt-Interaktionsanalysen (Purcell, 2002) zusammenfassend dargestellt, welche die Höhe von Erblichkeitsschätzungen über den Range des SES untersuchen. Die Ergebnisse dieser Studien liefern ein weiteres Beispiel für die Populationsabhängigkeit von Erblichkeitsschätzungen, denn insgesamt betrachtet wächst die Evidenz für eine variierende Erblichkeit kognitiver Fähigkeiten in Abhängigkeit der Höhe des SES (Nisbett et al., 2012). In einer frühen Studie finden Turkheimer, Haley, Waldron, D'Onofrio und Gottesman (2003) bei Familien mit niedrigem SES, dass nahezu die gesamte phänotypische Varianz in der Intelligenz durch Effekte der geteilten Umwelt erklärt werden kann, während am oberen Ende der SES Verteilung praktisch die gesamte Varianz durch genetische Effekte erklärt wird. Der gleiche Trend findet sich bei Harden, Turkheimer und Loehlin (2007), wenngleich die Ergebnisse dieser Studie auf Grund der relativ geringen Gesamtvarianz im SES nur eingeschränkt vergleichbar sind. In einer Längsschnittstudie mit Kindern, die im Alter von 10 Monaten zum ersten und im Alter von 2 Jahren zum zweiten Mal hinsichtlich ihrer kognitiven Entwicklung getestet wurden, zeigte sich zunächst keinerlei Unterschied in der Höhe der Erblichkeitsschätzung in Abhängigkeit des SES. Im Alter von 2 Jahren fanden sich jedoch bedeutsame Unterschiede für Kinder aus Familien mit niedrigem SES ($\approx 5\%$) und Kindern aus Familien mit hohem SES ($\approx 50\%$; Tucker-Drob, Rhemtulla, Harden, Turkheimer & Fask, 2011). Die aktuellsten Ergebnisse basieren auf Daten von 2-, 3-, 4-, 7-, 9-, 10-, 12- und 14-jährigen Kindern der TEDS Studie (Hanscombe et al., 2012). Die Autoren schließen aus den Ergebnissen, dass es in Familien mit niedrigem SES eine höhere Varianz geteilter Umwelteinflüsse gibt als in Familien mit hohem SES, wodurch der *relative* Varianzteil genetischer Effekte in Familien am unteren Ende der SES Verteilung niedriger ausfällt.

Das Befundmuster kann dennoch keineswegs als einheitlich bezeichnet werden. So finden sich auch Studien, in denen kein moderierender Effekt des SES gefunden wurde (Grant et al., 2010; van der Sluis, Willemsen, de Geus, Boomsma & Posthuma, 2008) oder ein Effekt in umgekehrter Richtung (höhere Erblichkeit am unteren Ende des SES)

vorlag (Asbury, Wachs & Plomin, 2005). Als Erklärung dieser divergierenden Befunde werden von Hanscombe et al. (2012) vor allem Unterschiede in der methodischen Umsetzung sowie Unterschiede in der Stichprobenzusammensetzung angeführt (vgl. auch Hanscombe et al., 2012, Tabelle 1, Seite 2).

4.6.3 Verhaltensgenetische Studien zur Motivation

Im Gegensatz zur Fülle an Befunden zur Erbllichkeit kognitiver Fähigkeiten haben sich bisher nur wenige Studien mit der Ätiologie motivationaler Merkmale beschäftigt. Dies mag u.a. der Tatsache geschuldet sein, dass klassische Motivationstheorien, z. B. die Theorie der Selbstwirksamkeit (Bandura, 1997) oder die Theorie des Selbstkonzeptes (Shavelson et al., 1976), von einer umweltbedingten Beeinflussung der Motivation ausgehen (Krapp, 2005).

Auch das Erwartungs-Wert-Modell von Eccles und Kollegen (vgl. Kapitel 3.3; Eccles et al., 1983; Eccles & Wigfield, 2002; Wigfield & Eccles, 2000) fokussiert auf eine umweltbedingte Beeinflussung von FSK und IM. Innerhalb des Modells wird davon ausgegangen, dass individuelle Unterschiede primär durch das Einwirken elterlicher Überzeugungen, Erwartungen, Einstellungen und Verhaltensweisen geformt werden (Eccles et al., 1983; Wigfield & Eccles, 2000). Diese Einflüsse werden implizit als geteilte Umwelteinflüsse verstanden, die auf Kinder, die in der gleichen Familie aufwachsen, gleichsam wirken und somit zu deren Ähnlichkeit beitragen. Entscheidend ist in diesem Kontext jedoch die Frage, ob es sich hierbei tatsächlich um einen Effekt der geteilten Umwelt handelt, oder ob vielmehr von Effekten der nichtgeteilten Umwelt ausgegangen werden muss, die zur Unähnlichkeit von Individuen, die in der gleichen Familie aufgewachsen sind, beitragen. Ein solcher Effekt wäre beispielsweise dann zu finden, wenn das Verhalten oder die Einstellung der Eltern unterschiedlich von den Kindern wahrgenommen wird, oder sie dieses unterschiedlich interpretieren (Harris, 1995). Andererseits ist es auch möglich, dass sich Eltern als Reaktion auf das jeweilige Temperament, Talent oder Verhalten der Kinder jedem Kind gegenüber unterschiedlich verhalten. Auch dies würde in Effekten der nichtgeteilten Umwelt resultieren (Plomin et al., 2001).

Innerhalb genetisch sensitiver Designs kann auch eine weitere Annahme des Eccles Modells geprüft werden: Bei der oben angesprochenen Beeinflussung der Kinder durch elterliche Überzeugungen, Erwartungen, Einstellungen und Verhaltensweisen wird im Modell eine mögliche Weitergabe von Genen der Parental- auf die Filialgeneration negiert. Dieser Annahme folgend sollten sich in verhaltensgenetischen Studien vor allem Effekte der Umwelt und allenfalls geringe genetische Effekte finden (F. M. Spinath et al., 2008a).

Studien zur Ätiologie der beiden Hauptkomponenten FSK und IM des Erwartungswert-Modells von Eccles und Kollegen (Eccles et al., 1983; Eccles & Wigfield, 2002; Wigfield & Eccles, 2000) liegen bisher ausschließlich aus der britischen Zwillingsstudie TEDS sowie der deutschen Studie KoSMoS vor. Durch deren korrespondierende Operationalisierung motivationaler Variablen ist es möglich, die Ergebnisse beider Studien direkt miteinander zu vergleichen.

Im Mittel liegt der genetische Einfluss auf die betrachteten motivationalen Variablen bei fachspezifischer Operationalisierung in der TEDS Stichprobe 9-jähriger Zwillinge ($N = 2.232$ Paare) sowohl für das FSK als auch die IM für Jungen und Mädchen bei ca. 40% (größtenteils in Form nicht-additiver genetischer Effekte). Die verbleibenden 60% der Varianz werden durch nichtgeteilte Umwelteffekte erklärt. Eine Ausnahme bildet für Jungen die IM im Fach Englisch, hier liegt der genetische Effekt mit 0.14 deutlich niedriger und es findet sich ein Effekt der geteilten Umwelt von 0.21 (F. M. Spinath et al., 2008a).

Auf Basis der gleichen Ausgangsstichprobe 9-jähriger untersuchen Greven, Harlaar, Kovas, Chamorro-Premuzic und Plomin (2009) ebenfalls das FSK, im Unterschied zu F. M. Spinath et al. (2008a) wird hier jedoch eine latente Modellierung auf globaler Ebene (als Konglomerat der Fähigkeitsselbsteinschätzung in Englisch, Mathematik und Naturwissenschaften) gewählt. Erwartungsgemäß findet sich wiederum ein moderater genetischer Effekt von 0.48, der Einfluss der Umwelt erklärt sich auch hier größtenteils durch nichtgeteilte Umwelteffekte ($e^2 = 0.43$, $c^2 = 0.09$). Durch die latente Modellierung kann dieser als reiner nichtgeteilter Umwelteffekt interpretiert werden (Greven et al., 2009).

Ebenfalls basierend auf TEDS Daten wird in einer aktuellen Studie von Y. L. L. Luo, Kovas, Haworth und Plomin (2011) eine globale Selbsteinschätzung im Fach Mathematik aus dem entsprechenden FSK und der IM generiert. Signifikante genetische Effekte liegen sowohl im Alter von 9 als auch im Alter von 12 Jahren vor (40% bzw. 43%), die restliche Varianz geht auf nichtgeteilte Umwelteffekte zurück.

Im Vergleich zu TEDS, in der das FSK und die IM lediglich zu zwei Zeitpunkten (im Alter von 9 und 12 Jahren) erfasst wurden (Oliver & Plomin, 2007), bietet die KoSMoS Studie, trotz einer weitaus geringeren Stichprobengröße den Vorteil, dass beide Motivationskomponenten bei jeder Messung erfasst wurden (vgl. Kapitel 8.4). Bisher liegen Ergebnisse der ersten und dritten Erhebungswelle vor, eine längsschnittliche Betrachtung über den gesamten Erhebungszeitraum steht allerdings noch aus.

Es lässt sich festhalten, dass das Muster einer substanziellen genetischen Beeinflussung und der Relevanz nichtgeteilter Umwelteffekte auch in KoSMoS Bestätigung findet. Für den ersten Erhebungszeitpunkt ($N = 407$ Paare im Alter zwischen 7 und 11 Jahren)

berichten F. M. Spinath et al. (2008b) einen über die Schulfächer Mathematik und Deutsch hinweg gemittelten genetischen Effekt von 0.55 für das FSK und 0.42 für die IM. Die verbleibende Varianz wurde jeweils durch Effekte der nichtgeteilten Umwelt aufgeklärt. Für das FSK im Alter von 13 Jahren ($N = 280$ Paare) liegt der genetische Effekt im Fach Deutsch mit 0.59 deutlich höher als im Fach Mathematik mit 0.35, in beiden Fällen wird die verbleibende Varianz vollständig durch die nichtgeteilte Umwelt erklärt (Gottschling et al., 2012). Dies unterstreicht zum einen die Wichtigkeit einer fachspezifischen Modellierung des FSKs, um mögliche Unterschiede in den Ursachen interindividueller Differenzen berücksichtigen zu können. Zum anderen stellt sich die Frage, worauf sich diese Unterschiede zurückführen lassen. Die Autoren führen hier potenziell vorliegende Kontrasteffekte für die Fähigkeitsselbsteinschätzung in Deutsch als mögliche Erklärung an.

Es kann an dieser Stelle zusammenfassend festgehalten werden, dass sich die Relevanz der berichteten Ergebnisse in Bezug auf das FSK und die IM in erster Linie aus der gefundenen Bedeutung nichtgeteilter Umwelteffekte und der relativen Bedeutungslosigkeit geteilter Umwelteffekte ergibt. Denn entgegen der theoretischen Modellannahmen finden sich zum einen sowohl in der englischen als auch in der deutschen Stichprobe substanzielle genetische Einflüsse auf FSK und IM von ca. 40%. Sie liegen damit in ähnlicher Höhe wie die genetische Beeinflussung kognitiver Fähigkeiten. Zum anderen hat sich in beiden Stichproben gezeigt, dass der Einfluss der Umwelt in erster Linie nichtgeteilter Art ist.

Neben Befunden zu den in vorliegender Arbeit betrachteten motivationalen Variablen liegen auch Ergebnisse zu anderen Motivationskonstrukten vor. Generell bestätigt sich sowohl für den globalen als auch domänenspezifischen *Selbstwert* in unterschiedlichen Altersbereichen ein substanzieller genetischer Einfluss und die Bedeutung der nichtgeteilten Umwelt (vgl. z. B. Kendler, Gardner & Prescott, 1998; Neiss, Sedikides & Stevenson, 2002; Roy, Neale & Kendler, 1995). Weiterhin weist auch das im Rahmen der *MTFS* erhobene *Selbstkonzept* eine moderate Erblichkeit von 30% auf, allerdings wurden hier lediglich Mädchen im Alter zwischen 11 und 13 Jahren betrachtet. Analog zu den zuvor berichteten Ergebnissen liegen auch in diesem Fall in erster Linie nichtgeteilte Umwelteffekte vor, während der Einfluss der geteilten Umwelt als unbedeutend eingestuft werden kann (Hur, McGue & Iacono, 1998).

4.6.4 Verhaltensgenetische Studien zur gemeinsamen Betrachtung schulischen Erfolgs und seiner Prädiktoren

Im Kontext der vorliegenden Arbeit sind auch multivariate verhaltensgenetische Ergebnisse zur Beziehung der Prädiktoren Motivation sowie allgemeine kognitive Fähigkeit und dem Kriterium Schulerfolg relevant. Nachfolgend werden daher Ergebnisse aus verhaltensgenetischen Studien berichtet, die diese Variablen simultan berücksichtigen. Allerdings liegen bisher lediglich Ergebnisse aus TEDS und KoSMoS vor, so dass die Generalisierbarkeit der Ergebnisse auf andere Altersbereiche und Länder eingeschränkt ist.

Greven und Kollegen analysieren in einem multivariaten genetischen Design 3.785 Zwillingspaare aus TEDS (Greven et al., 2009). Querschnittlich wurden Intelligenz, FSK und Schulleistung im Alter von 9 Jahren betrachtet, für die längsschnittliche Analyse wurde die kognitive Fähigkeit im Alter von 7 Jahren, das FSK im Alter von 9 Jahren und schließlich die Schulleistung im Alter von 10 Jahren herangezogen. Insgesamt fand sich eine bedeutsame genetische Überlappung der drei Konstrukte von über 50%, sowohl bei simultaner Erfassung der Konstrukte als auch bei Erfassung zu verschiedenen Zeitpunkten. Die Autoren interpretieren dies einerseits als Hinweis auf ein gemeinsames Set an Genen, welches alle Variablen beeinflusst und andererseits als Hinweis auf eine gewisse Stabilität der genetischen Faktoren (Greven et al., 2009). Ein geringer Teil der genetischen Effekte im FSK klärt zudem – unabhängig von Intelligenz – einen Teil der Varianz schulischer Leistung auf. Dies gilt analog wiederum für die längsschnittliche Betrachtung, was darauf hinweist, dass die Vorhersage der Leistung durch vorangegangenes FSK in erster Linie genetisch vermittelt ist (Greven et al., 2009).

Im Unterschied zu Greven et al. (2009) wurden in der entsprechenden multivariaten Analyse der KoSMoS Daten (basierend auf Daten der dritten Erhebungswelle, N=97 EZ und 183 ZZ Paare; mittleres Alter=13.1; SD=0.87) sowohl FSK als auch der Schulerfolg schulfachspezifisch für Mathematik und Deutsch modelliert (Gottschling et al., 2012). Auch in dieser Stichprobe findet die genetische Überlappung der drei Variablen Bestätigung, während für Umwelteinflüsse keine nennenswerte Überlappung festgestellt werden konnte. Übereinstimmend zu den Befunden aus TEDS wird der Großteil der Varianz im Schulerfolg durch solche genetischen Effekte erklärt, die auch auf kognitive Fähigkeiten wirken. Nur ein geringer Anteil an Varianz wird mit genetischen Effekten des FSKs geteilt (10% in Mathematik, 8% in Deutsch). Für die Effekte der nichtgeteilten Umwelt kann hingegen von einer hohen Spezifität für die einzelnen Messungen ausgegangen werden, was den Schluss nahelegt, dass es unterschiedliche nichtgeteilte Einflüsse sind, die auf Intelligenz, FSK und Schulerfolg wirken (Gottschling et al., 2012).

In der bereits erwähnten Studie von Y. L. L. Luo et al. (2011) wird die Selbsteinschätzung im Fach Mathematik (kombinierte Variable aus IM und FSK) und die *NC* Bewertung in Mathematik in einem genetischen CLP Design für das Alter 9 und 12 betrachtet. Hierbei wurden sowohl die Selbsteinschätzung als auch die *NC* Bewertung um Intelligenz kontrolliert. Die gefundenen CL Beziehungen zwischen Selbsteinschätzung und *NC* Bewertung waren für beide Beeinflussungsrichtungen (*Z1* Selbsteinschätzung auf *Z2 NC* Bewertung sowie *Z1 NC* Bewertung auf *Z2* Selbsteinschätzung) nahezu vollständig genetisch vermittelt ($> 95\%$). Einschränkend muss hier berücksichtigt werden, dass die CL Beziehungen mit 0.11 (*Z1* Selbsteinschätzung auf *Z2 NC* Bewertung) und 0.05 (*Z1 NC* Bewertung auf *Z2* Selbsteinschätzung) im Vergleich zu Ergebnissen anderer CL-Studien (z. B. Guay et al., 2003; Marsh et al., 1999; Marsh & Craven, 2006) eher niedrig ausfielen, ein Befund, der durch die Kontrolle von Intelligenz mitbedingt sein könnte. Als Folge sind die Reliabilitäten der Schätzungen der genetischen Effekte und der Umwelteffekte eingeschränkt was wiederum zu großen Konfidenzintervallen führt (Y. L. L. Luo et al., 2011). Die Interpretation der gefundenen Einflüsse auf die CL Beziehungen sollte daher nur unter Berücksichtigung dieses Umstandes erfolgen.

4.7 Abschließende Bewertung

Aus den vorausgehenden Ausführungen wird zunächst deutlich, dass die verhaltensgenetische Forschung eine wertvolle Ergänzung psychologischer Fragestellungen bietet. Durch den Einsatz genetisch sensitiver Designs wird es möglich, die verschiedenen Quellen phänotypischer Varianz innerhalb des menschlichen Verhaltensspektrums zu identifizieren. Hierzu bedient sich die quantitative Verhaltensgenetik unterschiedlicher Ansätze, so z. B. des Vergleichs gemeinsam aufgewachsener *EZ* und *ZZ*, der auch in der vorliegenden Arbeit Anwendung findet. Neben der univariaten Betrachtung einzelner Merkmale können zudem in multivariaten Modellierungen Quellen der Kovarianz zwischen Merkmalen, sowie Ursachen von Stabilität und Veränderung eruiert werden.

Ferner lässt sich aus den zuvor beschriebenen empirischen Befunden resümieren, dass der Kenntnisstand in Bezug auf die Erbllichkeit kognitiver Fähigkeiten als weitgehend gesichert bezeichnet werden kann. Ergebnisse zahlreicher Studien, die sich unterschiedlicher Designs bedienen und sich auf unterschiedliche Altersbereiche beziehen, belegen klare genetische Effekte in mittlerer Höhe, wobei die Höhe dieser Effekte von der frühen Kindheit bis zum hohen Erwachsenenalter linear ansteigt.

Die dargestellten Befunde machen allerdings auch deutlich, dass das Wissen um die Ursachen von interindividuellen Unterschieden in Schulerfolg (operationalisiert durch Fachnoten) und insbesondere hinsichtlich motivationaler Variablen weitaus weniger

stark vorangeschritten ist. Aber gerade in Bezug auf solche Variablen, die klassischerweise als durch die Umwelt beeinflusst angenommen werden, können verhaltensgenetische Studien als wichtige Informationsquelle herangezogen werden. Beispielsweise kann unter Rückgriff auf genetisch sensitive Designs die berichtete phänotypische interindividuelle Stabilität motivationaler Variablen (vgl. Kapitel 3.5) auf deren genetische und umweltbedingte Ursachen zurückgeführt werden. Hier wird gerade im Kontext einer Diskussion um die Beeinflussbarkeit motivationaler Voraussetzungen eine korrekte Interpretation der gefundenen Erblichkeitskoeffizienten salient, denn häufig wird in der öffentlichen Diskussion fälschlicherweise eine hohe Erblichkeit mit der Unveränderbarkeit eines Merkmals gleichgesetzt. Dass dieser Schluss so nicht gezogen werden darf, wurde in Abschnitt 4.2 ausführlich dargelegt.

Insgesamt fehlt es derzeit an Studien, die sich mit den Ursachen von Unterschieden vor allem in motivationalen Variablen befassen. Auch fehlt es bislang an längsschnittlichen verhaltensgenetischen Analysen, die die Ursachen von Stabilität und Veränderung schulischen Erfolgs auf der einen und motivationalen Merkmalen auf der anderen Seite beleuchten. Schließlich ist noch weitgehend ungeklärt, ob die phänotypisch gefundene reziproke Beziehung zwischen FSK und Schulerfolg durch genetische Effekte bestimmt wird oder durch eine umweltbedingte Beeinflussung erklärt werden kann. Für die Variable IM steht diese Betrachtung zudem noch gänzlich aus.

Abgeleitet aus den theoretischen Vorannahmen und empirischen Befunden werden im Folgenden die Fragestellungen und Hypothesen vorliegender Studie konkretisiert.

III. FRAGESTELLUNGEN UND HYPOTHESEN

Die vorliegende Arbeit fokussiert die längsschnittliche Vorhersage von schulischem Erfolg durch motivationale Faktoren unter Berücksichtigung allgemeiner kognitiver Fähigkeiten. Durch die Einbettung der Untersuchung in die KoSMoS Studie wird dabei neben einer rein phänotypischen auch die Ausweitung auf eine ätiologische Betrachtung möglich.

Basierend auf den in Teil II dargestellten theoretischen sowie empirischen Grundlagen werden im Folgenden die der Studie zugrunde liegenden Fragestellungen und Hypothesen hergeleitet. Der erste Teil der Arbeit bezieht sich auf die längsschnittliche, phänotypische Betrachtung der Vorhersage von SE durch die motivationalen Variablen FSK und IM unter Kontrolle allgemeiner kognitiver Fähigkeiten. Im zweiten, verhaltensgenetischen Teil der Arbeit, werden die Variablen hinsichtlich ihrer Ätiologie sowohl quer- als auch längsschnittlich betrachtet.

5 Fragestellung und Hypothesen der phänotypischen Betrachtung

Im ersten Teil der Untersuchung stehen zunächst Fragen nach der intra- und interindividuellen Stabilität in den Variablen SE, FSK, IM sowie allgemeiner kognitiver Fähigkeit im Mittelpunkt der Betrachtung. Diese können durch das längsschnittliche Design von KoSMoS mit drei Messzeitpunkten, die im Abstand von je zwei Jahren realisiert wurden, adäquat beantwortet werden. Zeitlich gesehen lag für alle teilnehmenden Kinder der Übergang von der Grund- in die Sekundarschule innerhalb dieses Untersuchungszeitraumes. Zudem bietet KoSMoS den Vorteil, dass die interessierenden motivationalen Variablen mit dem jeweils gleichen Messinstrument erfasst wurden, wodurch vorliegende Veränderungen über die Zeit auch als tatsächliche intra- und interindividuelle Veränderungen im FSK und der IM interpretiert werden können (Schmiedek & Lindenberger, 2007).

In einem sich anschließenden Schritt wird die Vorhersage schulischen Erfolgs durch Motivation (FSK bzw. IM) und allgemeine kognitive Fähigkeiten über den Untersuchungsverlauf hinweg betrachtet. Dabei liegt ein besonderer Fokus auf einer vergleichenden Betrachtung der Grundschul- und Sekundarschulzeit. Den Modellannahmen des Erwartungs-Wert-Modells (Eccles et al., 1983) sowie den Annahmen zur Struktur von Selbstkonzepten (z. B. Möller & Köller, 2004; Shavelson et al., 1976) folgend und unter Berücksichtigung der Befunde aus verschiedenen empirischen Studien (z. B. Möller & Trautwein, 2009; B. Spinath et al., 2006; H. S. Weber, 2012) werden die motivationalen Variablen fachspezifisch (Mathematik und Deutsch) operationalisiert. Wenngleich keine fachspezifischen Hypothesen formuliert werden, wird in der Interpretation der Befunde auf potenzielle Unterschiede zwischen den Fächern eingegangen werden.

5.1 Intra- und Interindividuelle Stabilität der Variablen

Im Rahmen längsschnittlicher Studien interessiert zunächst die Frage nach der intra- bzw. interindividuellen Stabilität der untersuchten Variablen über den Untersuchungszeitraum hinweg. Neben der Betrachtung von Mittelwertsveränderungen, speziell nach

dem Übergang in die Sekundarschule und der damit verbundenen Veränderung der Schulumwelt, ist auch die Frage nach einer Veränderung der Varianzen von Relevanz (Rietz & Rudinger, 2007). Ist die Varianz insgesamt eingeschränkt, wirkt sich dies auch auf die Höhe der zu erwartenden korrelativen Zusammenhänge aus (Amelang & Zielinski, 2002). Neben diesem methodischen Aspekt kommt auch der inhaltlichen Interpretation einer Zu- oder Abnahme der Variabilität in den betrachteten Variablen eine wichtige Bedeutung zu.

Bezogen auf die interindividuelle Stabilität steht die Frage im Mittelpunkt, ob die Kinder – unabhängig von einer eventuellen Mittelwertsveränderung – ihren relativen Rangplatz beibehalten. Liegt eine hohe interindividuelle Stabilität vor, kann davon ausgegangen werden, dass Merkmalsunterschiede zwischen den Kindern über die Zeit hinweg gleich bleiben (Schmitt, 2003).

Wie aus den Ausführungen in Teil II hervorgeht, liegen bereits umfassende Forschungsbefunde in Bezug auf die intra- und interindividuelle Stabilität schulischen Erfolgs sowie motivationaler Variablen und allgemeinen kognitiven Fähigkeiten vor.

Für den schulischen Erfolg kann dabei von einem kontinuierlichen Absinken über die Schullaufbahn hinweg bei steigender Varianz ausgegangen werden (z. B. Alspaugh, 1998; Ball et al., 2006). Ein besonders einschneidender Einbruch scheint sich dabei direkt nach dem Wechsel in das weiterführende Schulsystem (insbesondere bei Wechsel aufs Gymnasium) zu zeigen, was vielfach auf die Veränderungen in der Leistungsbeurteilung und einen Wechsel der Bezugsgruppe zurückgeführt wurde (z. B. Ball et al., 2006; Crockett et al., 1989; Tent, 2006). Auf Ebene von Merkmalsunterschieden kann hingegen von einer sehr hohen Stabilität schulischen Erfolgs ausgegangen werden (Tent, 2006; vgl. Kapitel 1.2). Analog zur Entwicklung schulischen Erfolgs findet sich auch in motivationalen Variablen ein negativer Trend über die Schullaufbahn hinweg (vgl. Kapitel 3.5). Dies wird unter anderem mit einer realistischer werdenden Einschätzung der eigenen Fähigkeiten sowie bestimmten Bezugsrahmeneffekten erklärt (z. B. Bouffard et al., 2003; Jacobs et al., 2002; D. Stipek & Iver, 1989). Dabei handelt es sich auf interindividueller Ebene betrachtet insbesondere beim FSK um ein sehr stabiles Merkmal (z. B. B. Spinath & F. M. Spinath, 2005b). Auch die Befundlage für allgemeine kognitive Fähigkeiten spricht eindeutig für eine hohe intra- und interindividuelle Stabilität in kognitiven Merkmalen ab dem mittleren Kindesalter (z. B. Moffitt et al., 1993; F. M. Spinath et al., 2003, vgl. Kapitel 2.4).

Hieraus abgeleitet werden folgende Hypothesen bezüglich der intra- und interindividuellen Stabilität schulischen Erfolgs und seiner Prädiktoren spezifiziert:

Hypothese 1a: Intraindividuelle Stabilität von SE, FSK, IM und allgemeinen kognitiven Fähigkeiten

Die Mittelwerte im schulischen Erfolg sowie im FSK und der IM nehmen über den Untersuchungsverlauf hinweg signifikant ab. Die Varianz in den Variablen steigt dabei leicht an.

Die Mittelwerte in der kognitiven Fähigkeit bleiben konstant.

Hypothese 1b: Interindividuelle Stabilität von SE, FSK, IM und allgemeinen kognitiven Fähigkeiten

SE, FSK, IM und allgemeine kognitive Fähigkeiten zeigen eine hohe interindividuelle Stabilität. Diese liegt für SE und FSK in einem sehr hohen Bereich, während für IM und allgemeine kognitive Fähigkeiten eine mittlere Stabilität vorliegt.

5.2 Längsschnittliche Analyse der Vorhersage von Schulerfolg

Kognitive Fähigkeiten stellen mit einer durchschnittlichen Varianzaufklärung von 25% den stärksten Einzelprädiktor schulischen Erfolgs dar (Neisser et al., 1996), weshalb der Einfluss allgemeiner kognitiver Fähigkeiten auch in der vorliegenden Studie berücksichtigt wird. Dennoch bleibt ein nicht unerheblicher Teil der Varianz im SE unaufgeklärt, was den Fokus auf nicht-kognitive Prädiktoren, beispielsweise motivationale Variablen, lenkt. Die Bedeutung, die diesen Merkmalen weithin zugesprochen wird, erklärt sich vor allem aus deren angenommener leichteren Beeinflussbarkeit (z. B. Steinmayr & B. Spinath, 2009), was auch die Chancen für eine erfolgreiche Intervention erhöht. Die sich hieraus ergebende praktische Relevanz dieser Variablen ist dabei offenkundig. In diesem Zusammenhang konnten verschiedene Studien bereits empirische Bestätigung für die Bedeutung motivationaler Variablen im Hinblick auf die Vorhersage schulischen Erfolgs liefern, wobei die Zusammenhänge mit zunehmendem Alter und zunehmender Fähigkeit in der Regel größer werden (z. B. Guay et al., 2003; Helmke, 1997; Marsh und Craven, 2006; Steinmayr und B. Spinath, 2009; vgl. Kapitel 3.4). Erste Studien, die mit adäquaten Analyseverfahren die inkrementelle Validität motivationaler Variablen über kognitive Fähigkeiten hinaus beurteilen, unterstreichen die Relevanz, insbesondere des FSKs, für den schulischen Erfolg (siehe 3.4.3). Für IM haben sich demgegenüber vor allem bei jüngeren Kindern geringere Zusammenhänge gezeigt. Um diese ersten Befunde weiter zu festigen bedarf es an dieser Stelle weiterer, vorzugsweise längsschnittlicher Studien mit deren Hilfe auch eine Abschätzung der Relevanz von Entwicklungs- und Kohorteneffekten möglich wird.

Der beschriebene Zusammenhang zwischen FSK und SE wurde auch unter einer ent-

wicklungsorientierten Perspektive vielfach in der Literatur diskutiert, wobei hier die Frage nach der Verursachungsrichtung im Mittelpunkt der Betrachtung stand (vgl. Kapitel 3.4.1). In den meisten dieser Studien konnte dabei ein Modell der reziproken Beeinflussung in unterschiedlichen Altersgruppen bestätigt werden (z. B. Guay et al., 2003). Andere Studien legen die Vermutung nahe, dass vor allem in der Grundschulzeit eher von einem *skill-development* Modell ausgegangen werden kann (z. B. Valentine et al., 2004). Im Hinblick auf IM konnten bisher keine reziproken Beziehungen bestätigt werden, allerdings liegen hierzu bisher nur sehr wenige Studien vor (Marsh et al., 2005b). Zudem wurde in vielen Studien keine fachspezifische Operationalisierung vorgenommen.

Basierend auf den in 3.4 dargestellten empirischen Befunden werden folgende Hypothesen konkretisiert:

Hypothese 2a: Vorhersage von SE durch allgemeine kognitive Fähigkeiten, FSK und IM über die Zeit

Der *g*-Faktor und die motivationalen Faktoren FSK und IM stellen wichtige Determinanten des schulischen Erfolgs dar und leisten einen wichtigen Beitrag zur Varianzaufklärung in den betrachteten Altersbereichen. Der *g*-Faktor ist der varianzstärkste Prädiktor. Das FSK ist ein stärkerer Prädiktor als die IM. Der Zusammenhang zwischen FSK bzw. IM und SE wird mit zunehmendem Alter größer.

Hypothese 2b: Kausale Prädominanz von FSK und SE

Vor allem in der Grundschulzeit beeinflusst der SE das nachfolgende FSK. Nach dem Übergang in das weiterführende Schulsystem findet sich eine reziproke Beeinflussung beider Faktoren.

Explorativ (2c): Kausale Prädominanz von IM und SE

Die Variable IM wird explorativ im Hinblick auf die Beeinflussungsrichtung untersucht.

Die Kontrolle des Einflusses der allgemeinen kognitiven Fähigkeit als stärkster Prädiktor schulischen Erfolgs ermöglicht es, den inkrementellen Beitrag motivationaler Variablen über die Zeit hinweg zu beurteilen. Es werden daher beziehungsweise auf die in Kapitel 3.4 dargestellten empirischen Befunde folgende Hypothesen formuliert:

Hypothese 2d: Inkrementelle Validität von FSK und IM

Die unter 2a angenommene Beziehung zwischen den motivationalen Faktoren und SE bleibt auch bei gleichzeitiger Berücksichtigung allgemeiner

kognitiver Fähigkeiten bestehen. Die Vorhersageleistung von FSK und IM schwächt sich jedoch ab.

Hypothese 2e: Kausale Prädominanz von FSK bzw. IM und SE unter Berücksichtigung allgemeiner kognitiver Fähigkeiten

Die gefundenen Zusammenhänge bzw. Entwicklungstrends hinsichtlich der kausalen Prädominanz motivationaler Variablen und SE bleiben bei Kontrolle kognitiver Fähigkeiten bestehen, nehmen aber insgesamt ab.

6 Fragestellung und Hypothesen der ätiologischen Betrachtung

6.1 Univariante Betrachtung

Die Frage nach den Ursachen interindividueller Unterschiede kann im Kontext vorliegender Arbeit vor allem hinsichtlich der motivationalen Variablen und des schulischen Erfolgs bisherige empirische Evidenzen ergänzen. Während eine genetische Beeinflussung kognitiver Fähigkeiten in unterschiedlichen Altersbereichen als gesichert gelten kann (z. B. Plomin & F. M. Spinath, 2004; F. M. Spinath et al., 2003) liegen bisher nur vereinzelt Studienergebnisse zu motivationalen Variablen vor. Diese deuten jedoch auch für diese, allgemein als durch die Umwelt beeinflussbar angenommenen Faktoren, auf substantielle genetische Effekte hin (z. B. Gottschling et al., 2012; Greven et al., 2009; F. M. Spinath et al., 2008b). Über quantitative Veränderungen in der Höhe der genetischen und Umwelteffekte kann bisher keine Aussage getroffen werden. Hinsichtlich einer genetischen Beeinflussung von schulischer Leistung liegen bisher vor allem Studien zu Schulleistungstests vor. Diese berichten mittlere genetische Effekte und mittlere nichtgeteilte Umwelteffekte. Werden Schulnoten betrachtet, liegen die berichteten genetischen Effekte in einem hohen Bereich (siehe Kapitel 4.6).

Basierend auf den Ergebnissen der vorliegenden empirischen Studien leiten sich folgende Hypothesen ab:

Hypothese 3a: Univariat verhaltensgenetische Betrachtung von SE

Schulnoten in Mathematik und Deutsch zeigen in univariaten verhaltensgenetischen Analysen Einflüsse additiver genetischer Effekte, sowie Effekte der geteilten und nichtgeteilten Umwelt auf die interindividuelle Merkmalsvarianz. Diese bleiben über den Untersuchungsverlauf hinweg vergleichsweise stabil.

Hypothese 3b: Univariat verhaltensgenetische Betrachtung von allgemeinen kognitiven Fähigkeiten

Allgemeine kognitive Fähigkeiten zeigen in univariaten verhaltensgeneti-

schen Analysen Einflüsse additiver genetischer Effekte, sowie Effekte der geteilten und nichtgeteilten Umwelt auf die interindividuelle Merkmalsvarianz. Die genetischen Einflüsse gewinnen mit steigendem Alter an Bedeutung, während die Bedeutung geteilter Umwelt abnimmt.

Hypothese 3c: Univariat verhaltensgenetische Betrachtung von FSK und IM

FSK und IM in Mathematik und Deutsch zeigen in univariaten verhaltensgenetischen Analysen Einflüsse additiver bzw. nicht-additiver genetischer Effekte, sowie Effekte der nichtgeteilten Umwelt auf die interindividuelle Merkmalsvarianz. Diese bleiben über den Untersuchungsverlauf hinweg in ähnlicher Höhe bestehen.

6.2 Multivariate Betrachtung

6.2.1 Ursachen interindividueller Stabilität

Die längsschnittliche Ausweitung der ätiologischen Betrachtung mittels multivariater genetischer Analysen ermöglicht die Abschätzung der Ursachen interindividueller Stabilität und Veränderung in den untersuchten Variablen. Hier zeigt sich im Bereich kognitiver Fähigkeiten, dass vor allem Gene zur Stabilität in diesen Merkmalen beitragen (siehe Kapitel 4.6). Für FSK und IM sowie Schulnoten stehen längsschnittliche verhaltensgenetische Studien bisher noch aus. Es wäre allerdings zu vermuten, dass gerade nach dem Wechsel in ein anderes Schulsystem und der damit einhergehenden Veränderung der Schulumwelt nichtgeteilte Umwelteffekte eher Veränderung in diesen Variablen erklären.

Hieraus abgeleitet werden folgende Hypothesen spezifiziert:

Hypothese 4a: Ursachen interindividueller Stabilität und Veränderung allgemeiner kognitiver Fähigkeiten

Additive genetische Effekte tragen zur Stabilität in allgemeinen kognitiven Fähigkeiten bei.

Explorativ (4b): Ursachen interindividueller Stabilität und Veränderung in SE, FSK und IM

Die Ursachen interindividueller Stabilität und Veränderung werden explorativ untersucht. Es wird jedoch vermutet, dass vor allem nach dem Über-

gang in die weiterführende Schule nichtgeteilte Umwelteffekte Veränderung erklären.

6.2.2 Verhaltensgenetische cross-lagged Betrachtung der Vorhersage von SE

Für die unter Kapitel 5 formulierten Hypothesen der kausalen Prädominanz kann unter Rückgriff auf genetische *cross-lagged* Analysen ergänzend zur phänotypischen Betrachtung auch die Ätiologie der jeweiligen Kreuzpfade untersucht werden. Eine potenzielle genetische Vermittlung der Beziehung zwischen FSK bzw. IM und SE zu zwei Zeitpunkten wurde bisher lediglich in einer Studie untersucht. Diese findet für beide Beeinflussungsrichtungen eine sehr hohe genetische Vermittlung des Kreuzpfades (Y. L. L. Luo et al., 2011), wobei in dieser Studie FSK und IM als Globalfaktor modelliert wurden (vgl. Kapitel 4.6).

Basierend auf dieser Befundlage werden folgende Hypothesen spezifiziert:

Hypothese 5a: Verhaltensgenetische Analyse der längsschnittlichen Beziehung zwischen FSK und SE

Die kausale Beeinflussung von SE durch FSK sowie von FSK durch SE wird sowohl in Mathematik als auch in Deutsch primär genetisch vermittelt.

Hypothese 5b: Verhaltensgenetische Analyse der längsschnittlichen Beziehung zwischen IM und SE

Die kausale Beeinflussung von SE durch IM sowie von IM durch SE wird sowohl in Mathematik als auch in Deutsch primär genetisch vermittelt.

IV. DATENGRUNDLAGE UND METHODISCHES VORGEHEN

Der methodische Teil der vorliegenden Arbeit ist in vier Teile gegliedert: Zunächst wird die Zwillingsstudie KoSMoS und das Vorgehen bei der Auswahl der hier verwendeten Stichprobe zur Umsetzung des Multi-Kohorten-Sequenz-Designs, sowie der ätiologischen Analysen beschrieben. Hiernach werden die zur Operationalisierung eingesetzten Verfahren erläutert. In den beiden anschließenden Kapiteln werden die eingesetzten Analysen der phänotypischen und ätiologischen Betrachtung vorgestellt.

7 Stichprobe

Die Beantwortung der unter III formulierten Fragestellungen erfolgte aufbauend auf Daten der längsschnittlich angelegten, deutschen *Zwillingsstudie zu Einflüssen von kognitiven Fähigkeiten und selbst eingeschätzter Motivation auf Schulerfolg* (F. M. Spinath & Wolf, 2006).

Im Folgenden wird zunächst ein Überblick über den allgemeinen Aufbau und den Ablauf der Studie, sowie die Verankerung der vorliegenden Arbeit in den Kontext von KoSMoS gegeben. Daran schließt sich eine detaillierte Darstellung der verwendeten Stichprobe sowie der eingesetzten Instrumente an.

7.1 Die Zwillingsstudie zu Einflüssen von kognitiven Fähigkeiten und selbst eingeschätzter Motivation auf Schulerfolg (KoSMoS)

Die längsschnittlich konzipierte Zwillingsstudie KoSMoS wurde im Jahr 2005 in Anlehnung an die britische Zwillingsstudie TEDS an der Universität des Saarlandes initiiert und umfasst insgesamt drei Erhebungswellen, die jeweils im Abstand von zwei Jahren durchgeführt wurden. Ziel war die längsschnittliche Untersuchung von potenziellen individuellen sowie umweltbedingten Einflussfaktoren auf die schulische Leistungsentwicklung im deutschsprachigen Raum. Durch die Konzeption der Studie im genetisch sensitiven Design sind zudem Aussagen über die Ursachen von interindividuellen Unterschieden in diesen Variablen möglich.

Um eine differenzierte Betrachtung potenzieller Einflussfaktoren auf die schulische Leistungsentwicklung zu ermöglichen, wurde ein umfangreiches Set an individuellen, familiären und schulischen Variablen in Selbst- und zum Teil in Fremdeinschätzung erfasst (F. M. Spinath & Wolf, 2006).

Auf Seiten individueller Variablen der Kinder wurden unter anderem verbale und nonverbale kognitive Fähigkeiten, verschiedene Motivationskonstrukte, Persönlichkeit, wahrgenommener Erziehungsstil und wahrgenommenes Involvement der Eltern erfragt.

Als familiäre Variablen wurden neben anderen die Persönlichkeit der Eltern, ihre Wertvorstellungen, das Erziehungsverhalten und die elterliche Unterstützung erhoben. Ferner wurden demografische Angaben der Familie, der Alltag der Zwillinge, die Zygote und allgemeine Angaben zum Lern- und Leistungsumfeld der Zwillinge erfasst. Zur Operationalisierung des Schulerfolges wurden die Zensuren in unterschiedlichen Schulfächern abgefragt. Die Schulnoten und der besuchte Schultyp wurden für den gesamten Erhebungszeitraum von 2005 bis 2009 zum Teil retrospektiv telefonisch oder per Onlinebogen erhoben, wodurch Schulnoten für jedes Schuljahr seit dem Halbjahreszeugnis 2004/05 bis zum Halbjahreszeugnis 2009/10 vorliegen.

Bezüglich der Erhebungsmethode fand im Studienverlauf (von Messzeitpunkt 1 auf 2) eine Umstellung von einer Paper-Pencil Erhebung auf eine webbasierte Computertestung der Teilnehmer statt (eine ausführliche Beschreibung des Ablaufs der Onlinetestung findet sich bei Hegewald, 2009). Dieses Vorgehen stellt eine ökonomischere Alternative zur Paper-Pencil-Testung dar, die Vergleichbarkeit mit Daten aus Paper-Pencil-Erhebungen hinsichtlich Maßen interner Konsistenz und Skalenwerten kann mittlerweile als belegt gelten (z. B. Cole, Bedeian & Feild, 2006; Horswill & Coster, 2001). Im Vergleich zur Paper-Pencil-Erfassung findet sich in Online-Erhebungen eine weitaus geringere Anzahl an fehlenden Werten (z. B. Wood, Nosko, Desmarais, Ross & Irvine, 2006) und es kann im Allgemeinen mit einer geringer ausfallenden Antworttendenz in Richtung sozialer Erwünschtheit gerechnet werden (Joinson, 1999).

Auch in Bezug auf die kognitiven Fähigkeiten, die insgesamt zweimal (2005 und 2009) erfasst wurden, fand eine Umstellung der Erhebungsmethode im Studienverlauf statt: Während 2005 eine elternadministrierte Form der Testung umgesetzt wurde, wurden die Kinder im Jahr 2009 von geschulten Testleitern per Telefon getestet (zur näheren Beschreibung beider Vorgehensweisen siehe Kapitel 8). Eine Überprüfung der Äquivalenz der Messungen aus unterschiedlichen Erhebungszeitpunkten und unterschiedlichen Operationalisierungen wird in Kapitel 12.2 vorgenommen.

Eine Besonderheit der Studie ist ferner in der umgesetzten Rekrutierungsmethode zu sehen. Um eine selektive Vorauswahl hinsichtlich besonders motivierter Kinder bzw. Eltern zu vermeiden,¹ wurde auf eine Anwerbung über Medien und Zwillingclubs, wie es in Ländern ohne Zwillingregister gängige Praxis ist, verzichtet. Stattdessen wurde eine innovative Rekrutierungsmethode in Zusammenarbeit mit Meldeämtern aus Nordrhein-Westfalen und Thüringen realisiert (F. M. Spinath & Wolf, 2006). Zunächst wurden aus den verfügbaren Meldeamtsregistern sämtliche Daten von Personen mit gleichem Nachnamen, Geburtstag und Geburtsort gesammelt. Aus den hieraus resultierenden potenziellen Zwillingspaaren wurden zunächst die Geburtsjahrgänge 1995 bis

¹ In Zwillingstudien kommt es infolge dieser selektiven Auswahl häufig zu einer Überrepräsentation eineiiger Zwillinge.

1998 als interessierender Altersbereich ausgewählt und die Adressen mit Daten aus dem Telefonregister abgeglichen.² In der Folge wurden 715 Familien telefonisch und 1.190 Familien postalisch in Form eines Informationsschreibens über die Studie informiert. Ein Vergleich der beiden Vorgehensweisen spricht hinsichtlich der Teilnahmebereitschaft für eine Überlegenheit des telefonischen Erstkontaktes: Hier erklärten sich knapp zwei Drittel aller Familien zu einer Teilnahme bereit, im Gegensatz zu lediglich 26% nach postalischer Kontaktaufnahme (F. M. Spinath & Wolf, 2006). Anschließend wurden die Fragebogensets, bestehend aus einem Elternbogen, zwei Kinderfragebögen sowie dem Booklet der kognitiven Testung, einer Einverständniserklärung und einem frankierten Rückumschlag zu den Familien verschickt.

Für den ersten Erhebungszeitpunkt (Z1) im Jahr 2005 lagen schließlich insgesamt 408 vollständige Datensets (ausgefüllter Elternbogen sowie ausgefüllte Bögen von beiden Zwillingskindern) vor, was einer Rücklaufquote von ca. 46% entspricht. In der Umsetzung des angestrebten längsschnittlichen Studiendesigns wurden zwei weitere Messzeitpunkte in den Jahren 2007 (Z2; $N = 267$ Familien) und 2009 (Z3; $N = 209$ Familien) realisiert. Die Ausfallquote war mit 34.6% (2005 nach 2007) bzw. 48.8% (2005 nach 2009) insgesamt vergleichbar mit derjenigen von anderen großen Längsschnittstudien.^{3,4} Über den gesamten Erhebungszeitraum betrachtet nahmen von den ursprünglichen 816 Kindern 575 an mindestens zwei der drei Messzeitpunkte teil. Eine eingehende Beschreibung der Auswahl der hier verwendeten Untersuchungsstichprobe wird nachfolgend vorgenommen.

7.2 Auswahl der Untersuchungsstichprobe

Zur Beantwortung der Fragestellungen der vorliegenden Arbeit war es zunächst erforderlich, eine geeignete Untersuchungsstichprobe zusammenzustellen. Hierfür wurde die Stichprobe des ersten Erhebungszeitpunktes 2005 mit Daten von insgesamt 408 Familien zugrunde gelegt.

Wie aus Abbildung 7.1 hervorgeht wurden die Kinder entsprechend ihrer im Jahr 2005 besuchten Klassenstufe in fünf Kohorten eingeteilt. Als Untersuchungsstichprobe

² Die Anzahl falsch positiver Kontakte (keine Zwillinge bei gleichem Geburtsnamen, -tag und -ort) war dabei mit 2,4% sehr niedrig.

³ Z. B. der *Panel Analysis of Intimate Relationships and Family Dynamics*: Hier werden Ausfallquoten zwischen Welle 1 (2008/09) und Welle 3 (2010/2011) zwischen ca. 28% in Kohorte 1 (Jahrgang 1991-93) und ca. 43% in Kohorte 2 (Jahrgang 1981-83) berichtet. <http://www.pairfam.uni-bremen.de/de/daten/stichproben.html>.

⁴ Z. B. die *Bielefeld Longitudinal Study of Adult Twins*: Hier lag die Ausfallquote mit ca. 29% zwischen 1993 und 1994 noch in vergleichbarer Höhe, stieg aber zwischen 1993 und 1999 auf ca. 70% an. <http://www.uni-bielefeld.de/psychologie/ae/AE04/Forschung/index.html>.

(K_AS) wurden daraufhin diejenigen Kinder ausgewählt, die Kohorte 1, 2 oder 3 zugeordnet werden konnten, um ausreichend große Substichproben zu gewährleisten (im Folgenden als K1, K2 und K3 bezeichnet).

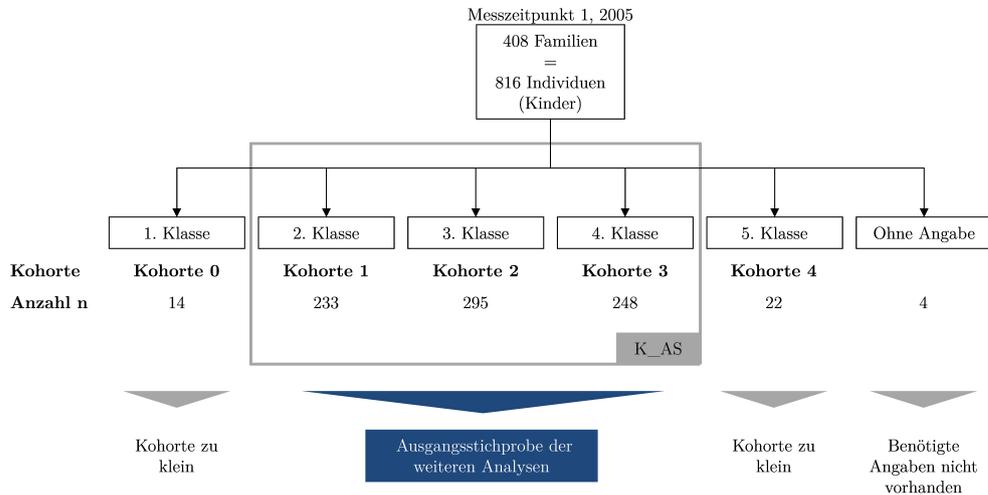


Abbildung 7.1. Beschreibung der Gesamtstichprobe, eingeteilt in 5 Kohorten (K0 bis K4).

Tabelle 7.1 zeigt die Zusammensetzung der Untersuchungsstichprobe K_AS des ersten Messzeitpunktes hinsichtlich Alter, Geschlecht, besuchter Klassenstufe sowie Zygote (eine detaillierte Aufschlüsselung über den gesamten Erhebungszeitraum hinweg findet sich in Anhang B). Kohorte 2 gehörten mit $n = 295$ die meisten Kinder an. Das Geschlechterverhältnis stellte sich in allen Kohorten als nahezu ausgeglichen dar. Das durchschnittliche Alter der Kinder lag zwischen 8.56 Jahren in Kohorte 1 und 10.12 Jahren in Kohorte 3.

Zur Bestimmung der Anzahl an EZ und ZZ Paaren wurden zunächst diejenigen Paare ausgeschlossen, die bei der ersten Erhebung in unterschiedlichen Klassenstufen waren ($n=23$ Paare), um eine eindeutige Zuordnung der Paare in Kohorten sicherzustellen. Außerdem wurden weitere acht Paare ausgeschlossen, deren Zygote nicht eindeutig bestimmt werden konnte (zur Zygotebestimmung vgl. Abschnitt 8.1). Dementsprechend lagen in der Untersuchungsstichprobe zum ersten Messzeitpunkt Daten von 361 Zwillingspaaren vor, die in etwa der Verteilung von Zwillingspaaren in der Population (1/3 EZ-Paare zu 2/3 ZZ-Paare) folgen.

Abbildung 7.2 informiert über die Teilnahme der Kinder der Untersuchungsstichprobe K_AS über den Untersuchungszeitraum von fünf Jahren hinweg. Die Nomenklatur der entsprechenden Teilstichproben zu den jeweiligen Zeitpunkten kann ebenfalls Abbildung 7.2 entnommen werden. Wie aus dieser hervorgeht, nahmen an der zweiten Erhebung im Jahr 2007 noch 508 Kinder ($N = 254$ Familien) und im Jahr 2009 noch

Tabelle 7.1
Beschreibung der Untersuchungsstichprobe K_AS zu Messzeitpunkt 1

	Klasse	N	Alter		Geschlecht	
			M	SD	m	w
Kohorte 1	2	233	8.56	0.54	116	117
Kohorte 2	3	295	9.52	0.53	152	143
Kohorte 3	4	248	10.12	0.64	115	133

Anmerkungen. M = Mittelwert; SD = Standardabweichung; m = männlich; w = weiblich.

	Anzahl EZ			Anzahl ZZ			
	m	w	Gesamt	m	w	gg	Gesamt
Kohorte 1	21	22	43	14	15	35	64
Kohorte 2	25	17	42	24	26	43	93
Kohorte 3	24	20	44	13	24	38	75
Gesamt			129				232

Anmerkungen. EZ = Eineiige Zwillinge; ZZ = Zweieiige Zwillinge; m = männlich; w = weiblich; gg = gegengeschlechtlich.

392 Kinder ($N = 196$ Familien) an KoSMoS teil, was einer Rücklaufquote von 65,5% respektive 50,3% der Untersuchungsstichprobe K_AS entspricht. Die Drop-out Quote ist damit insgesamt vergleichbar mit der anderer Schullängsschnittstudien (siehe auch Abschnitt 7.1). Die 575 Kinder, die an mindestens zwei der drei Erhebungen teilnahmen (K_AS_2), stellten die Stichprobe für die längsschnittlichen Analysen dar.

Wie zuvor bereits beschrieben lag in der Studie eine nicht unerhebliche Ausfallquote der Teilnehmer vor. Zur Überprüfung eines systematischen Ausfalls wurde eine Drop-out Analyse durchgeführt. Hierzu wurde die Gruppe der *Teilnehmer* mit der Gruppe der *Abbrecher* hinsichtlich der Variablen *Alter*, *Geschlecht*, *Zygotie* sowie dem *SES* verglichen. Die Ergebnisse der Analyse werden nachfolgend dargestellt, die berichteten Kennwerte beziehen sich auf eine reduzierte Stichprobe mit lediglich einem zufällig ausgewählten Zwilling pro Paar.

7.2.1 Drop-out-Analyse

Innerhalb längsschnittlicher Untersuchungsdesigns liegt in der Regel ein gewisser Ausfall der Studienteilnehmer vor, was allgemein als *Mortalität* der Stichprobe bezeichnet wird. Wie in Kapitel 9.1 beschrieben wird, können fehlende Daten mit Hilfe geeigneter Schätzalgorithmen adäquat ersetzt werden, so dass der Stichprobenausfall nicht

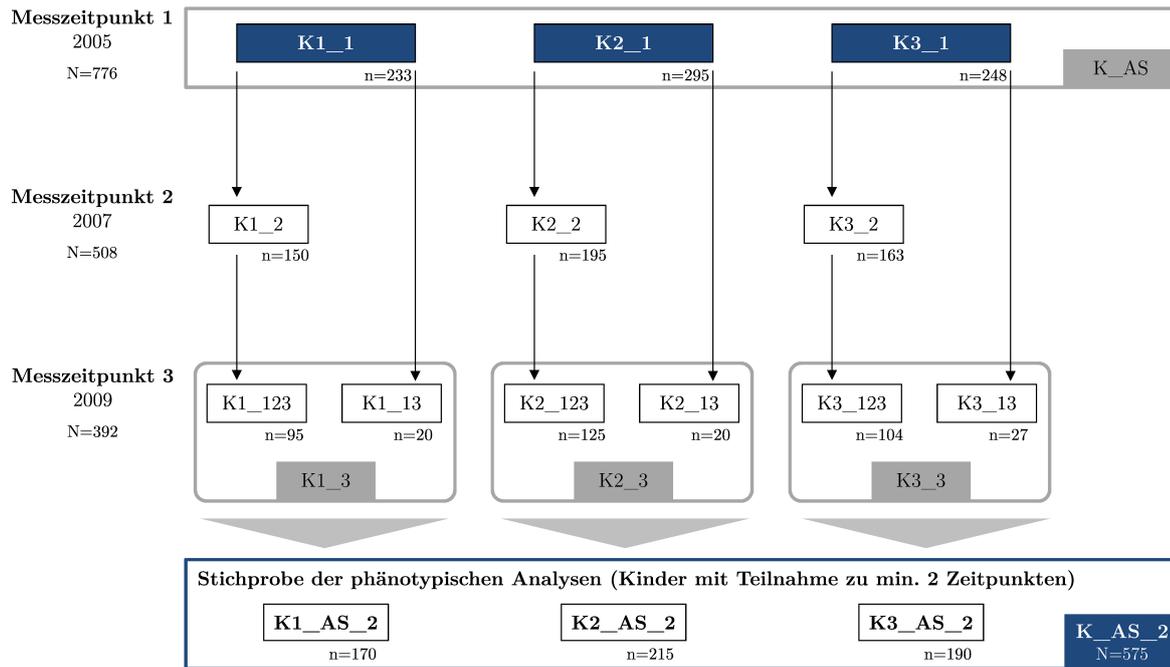


Abbildung 7.2. Verbleib der Kinder, gesplittet in Kohorten, über den Untersuchungsverlauf hinweg.

zwangsläufig zu einer eingeschränkten Aussagekraft der Ergebnisse führt. Dennoch ist es in einem vorgelagerten Schritt sinnvoll, den Drop-out in Bezug auf einen eventuell vorliegenden systematischen Ausfall hin zu untersuchen (Rietz & Rudinger, 2007). Auf diese Weise kann geklärt werden, ob sich die weiterhin an der Studie teilnehmenden Personen in bestimmten Merkmalen von den Abbrechern unterscheiden. Im Kontext der vorliegenden Arbeit werden hierfür als in Frage kommende Variablen das Alter der Kinder, ihr Geschlecht, ihre Zygote und schließlich der SES der Familie einer näheren Betrachtung unterzogen. Es wird geprüft, ob ein systematischer Ausfall für den zweiten und dritten Messzeitpunkt vorlag.

Bezüglich des *Alters* der Kinder kann festgehalten werden, dass es weder in der zweiten noch der dritten Erhebung einen signifikanten Unterschied zwischen der Gruppe der Teilnehmer (zu Z1: $M = 9.08$, $SD = 0.82$) und der Gruppe der Abbrecher (zu Z1: $M = 9.10$, $SD = 0.79$) gab ($t(404) = 0.18$, $p = 0.86$; bzw. $t(404) = -0.24$, $p = 0.81$). Gleiches galt für *Geschlecht* (Z2: $\chi^2_{(1,n=142)} = 0.00$, $p = 0.97$ bzw. zu Z3: $\chi^2_{(1,n=203)} = 0.00$, $p = 1.00$) und *Zygote* (Z2: $\chi^2_{(5,n=142)} = 4.34$, $p = 0.50$ bzw. zu Z3: $\chi^2_{(5,n=203)} = 4.46$, $p = 0.49$). Es nahmen dementsprechend nicht mehr Jungen als Mädchen weiterhin an der Erhebung teil, noch lag eine systematische Verzerrung hinsichtlich einer bestimmten Zygotegruppe vor.

Eine Einschränkung bzgl. der Vergleichbarkeit der Teilnehmer und Abbrecher fand sich allerdings im Hinblick auf den SES. Hier zeigte sich zum zweiten Erhebungszeitpunkt

ein höherer SES der Teilnehmer ($M = 5.58$, $SD = 2.17$) im Vergleich zu den Abbrechern ($M = 4.96$, $SD = 2.35$; $t(404) = -2.63$, $p = 0.01$). Auch in der dritten Erhebung zeigte sich dieser Effekt, allerdings in etwas geringerem Ausmaß (Teilnehmer: $M = 5.59$, $SD = 2.25$; Abbrecher: $M = 5.14$, $SD = 2.23$; $t(404) = -2.01$, $p = 0.05$).

Um diesen Aspekt näher zu beleuchten, wurde zusätzlich die Untersuchungsstichprobe K_AS hinsichtlich einer positiven Selektion in Bezug auf den familiären SES (als Indikator wurde der höchste elterliche Bildungsabschluss herangezogen) im Vergleich zu einer repräsentativen gesamtdeutschen Stichprobe⁵ betrachtet.

Anzahl Eltern nach höchstem Bildungsabschluss Prozent

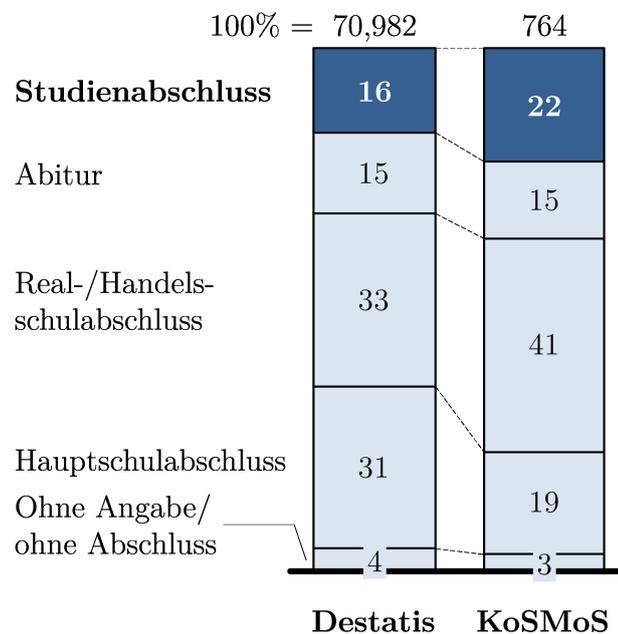


Abbildung 7.3. Höchster Bildungsabschluss der Familie (gemittelter Wert beider Eltern) im Vergleich mit einer repräsentativen deutschen Vergleichsstichprobe.

Hieraus lässt sich schlussfolgern, dass eine Überrepräsentation mittlerer und hoher Bildungsabschlüsse und eine Unterrepräsentation niedriger Bildungsabschlüsse in KoSMoS vorlag. Generell erklärten sich folglich eher Eltern höherer Bildungsschichten zur Teilnahme an der Studie bereit (vgl. Abbildung 7.3). In KoSMoS lag der Anteil an Eltern mit Studienabschluss 6% über dem Bundesdurchschnitt. Hiermit vergleichbar ist der Unterschied bei Real-/und Handelsschulabschlüssen, der in KoSMoS um 8% höher

⁵ Daten basieren auf dem Mikrozensus 2010 (Statistisches Bundesamt, 2010). Für den Vergleich wurde aus der gesamtdeutschen Stichprobe eine Stichprobe gezogen, die hinsichtlich der Altersverteilung (Eltern) der KoSMoS Stichprobe entsprach.

ausfiel als im Bundesvergleich. Die größte Differenz findet sich bei der Häufigkeit von Hauptschulabschlüssen: In der gesamtdeutschen Referenzgruppe hat knapp ein Drittel der Befragten einen Hauptschulabschluss, während dies in der KoSMoS Stichprobe lediglich bei 19% der Fall war.

Diese Beobachtung einer vorliegenden Selektion im Hinblick auf höhere Bildungsabschlüsse der Eltern in der KoSMoS Stichprobe setzt sich auch bei der Verteilung der Schulformen der Kinder zu Z3 (hier besuchten alle Kinder die Sekundarschule) fort. Wie aus Abbildung 7.4 klar hervorgeht, lag der Anteil an Kindern die das Gymnasium besuchen in der KoSMoS Stichprobe fast doppelt so hoch wie in einer bundesdeutschen Vergleichsstichprobe⁶ (35% vs. 64%). Dementgegen waren nur 4% der Kinder der Untersuchungsstichprobe auf der Hauptschule (vs. 16% im Bundesdurchschnitt). Dieser Selektivität ist bei der Interpretation der Ergebnisse der vorliegenden Studie entsprechend Rechnung zu tragen.

Anzahl Kinder (3. Messzeitpunkt) je Schulform

Prozent

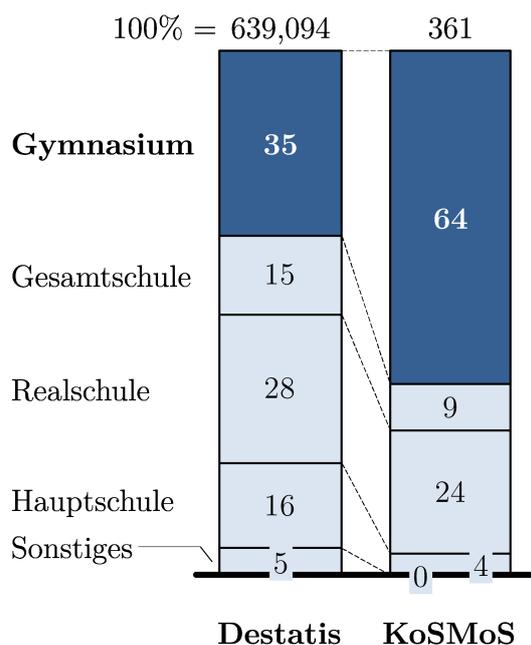


Abbildung 7.4. Besuchte Schulform der Kinder im Vergleich mit einer deutschen Vergleichsstichprobe.

Nach dieser allgemeinen Beschreibung der Auswahl einer geeigneten Untersuchungsstichprobe werden im Folgenden die Analysestichproben, die zur Beantwortung der phänotypischen und ätiologischen Fragestellungen herangezogen wurden, beschrieben.

⁶ Basierend auf Daten aus dem Mikrozensus 2010 wurde eine Vergleichsstichprobe besuchter Schultypen im Schuljahr 2008/09 in Nordrhein-Westfalen und Thüringen gezogen, weil die KoSMoS Stichprobe in diesen Ländern rekrutiert wurde (Statistisches Bundesamt, 2010).

7.2.2 Stichproben der phänotypischen und ätiologischen Analysen

Wie aus Kapitel 5 und 6 hervorgeht werden phänotypisch längsschnittliche Fragestellungen in Bezug auf intra- und interindividuelle Stabilität des Schulerfolges und seiner Prädiktoren untersucht. Die verhaltensgenetischen Analysen beziehen sich hingegen sowohl auf eine querschnittliche als auch eine längsschnittliche Betrachtung. Die hierfür ausgewählten Analysestichproben setzen sich wie nachfolgend beschrieben zusammen.

Stichprobe zur Beantwortung der phänotypischen Fragestellungen

Wie im vorherigen Abschnitt bereits angesprochen, wurden zur Beantwortung der Fragestellungen der phänotypischen Betrachtung diejenigen Kinder als Stichprobe zugrunde gelegt, die an mindestens zwei der drei Erhebungszeitpunkte teilgenommen haben (K_AS_2, $N = 575$; vgl. Abbildung 7.2). Die fehlenden Daten werden bei Bestätigung eines zufälligen Ausfalls (vgl. hierzu Kapitel 9) aus den vorliegenden Daten unter Rückgriff auf eine Maximum-Likelihood-Schätzung ersetzt. Die Überprüfung dieser Voraussetzung wird in Abschnitt 12.1 vorgenommen.

Tabelle 7.2 enthält eine Übersicht über K_AS_2 zum ersten Messzeitpunkt hinsichtlich Alter, Geschlecht und besuchter Klassenstufe, gesplittet in Kohorte 1 bis Kohorte 3 (nachfolgend als K1_AS_2, K2_AS_2 und K3_AS_2 bezeichnet).

Tabelle 7.2

Beschreibung der Analysestichprobe der phänotypischen Analysen zu Messzeitpunkt 1

	Klasse	N	Alter		Geschlecht	
			M	SD	m	w
K1_AS_2	2	170	7.62	0.54	81	89
K2_AS_2	3	215	8.53	0.52	111	104
K3_AS_2	4	190	9.30	0.50	84	106

Anmerkungen. M = Mittelwert; SD = Standardabweichung; m = männlich; w = weiblich.

Kohorte 2 stellt dementsprechend die stärkste Gruppe dar, das Verhältnis von Jungen und Mädchen war lediglich in Kohorte 3 leicht zu Gunsten der Mädchen verschoben.

Um Rückschlüsse auf einen möglichen Einfluss des Sekundarschulübergangs auf die Prädiktion schulischen Erfolgs durch die betrachteten Determinanten ziehen zu können, ist für die längsschnittliche Betrachtung weiterhin die besuchte Klassenstufe der drei Kohorten zu den verschiedenen Untersuchungszeitpunkten von Relevanz. Diese sind

Abbildung 7.5 zu entnehmen. Für Kohorte 1 lag der Übergang in die weiterführende Schule entsprechend zwischen $Z2$ und $Z3$, für Kohorten 2 und 3 hingegen zwischen $Z1$ und $Z2$.

		Z 1		Z 2		Z 3	
Schuljahr		2004/05	2005/06	2006/07	2007/08	2008/09	2009/10
Klassenstufe	Kohorte 1	2. Klasse K1_1		4. Klasse K1_2		6. Klasse K1_3	
	Kohorte 2	3. Klasse K2_1		5. Klasse K2_2		7. Klasse K2_3	
	Kohorte 3	4. Klasse K3_1		6. Klasse K3_2		8. Klasse K3_3	

Abbildung 7.5. Klassenstufe der Kinder, gesplittet in Kohorten, über den Untersuchungsverlauf hinweg. Kohorte 1 wechselte zwischen $Z2$ und $Z3$ in die Sekundarschule, Kohorte 2 und Kohorte 3 zwischen $Z1$ und $Z2$.

Stichprobe zur Beantwortung der ätiologischen Fragestellungen

Im Rahmen der verhaltensgenetischen Analysen werden zunächst potenzielle quantitative und qualitative Unterschiede in den Ursachen interindividueller Unterschiede in den Variablen SE, Motivation und allgemeinen kognitiven Fähigkeiten in verschiedenen Altersbereichen untersucht. Zu diesem Zweck wird die Erbllichkeit der Variablen in den unterschiedlichen Altersbereichen der drei Messzeitpunkte betrachtet (Abbildung 7.5). Hierfür bedarf es keiner längsschnittlichen Stichprobe und auch der Zeitpunkt des Übergangs in das weiterführende Schulsystem ist in diesem Zusammenhang nicht von Bedeutung. Vielmehr kann querschnittlich auf sämtliche Daten zurückgegriffen werden, die zu den entsprechenden Zeitpunkten vorlagen, um die größtmögliche Power der Analysestichprobe zu erreichen (zu $Z1$ wurde hierfür K_AS und für $Z2$ bzw. $Z3$ K_AS_2 verwendet). Die sich ergebende Anzahl an zur Verfügung stehenden Paaren kann Tabelle 7.3 entnommen werden.

Die längsschnittliche verhaltensgenetische Betrachtung zielt einerseits auf die Erklärung der Ursachen von Stabilität und Veränderung in den betrachteten Variablen. Andererseits wird die Frage der kausalen Prädominanz von Motivation und Schulerfolg aus einer ätiologischen Perspektive heraus betrachtet. Als Stichprobe wird hierfür analog zur phänotypischen Betrachtung K_AS_2 zugrunde gelegt. Um auch in diesem Fall eine größtmögliche Power der Stichprobe zu erreichen, werden für die multivariaten Analysen Daten von Kindern aus Kohorte 2 und 3 zusammengefasst, da diese

zur gleichen Zeit innerhalb des Untersuchungszeitraumes von der Grundschule in das weiterführende Schulsystem wechseln. Aufgrund der zu geringen Anzahl an Zwillingspaaren in Kohorte 1 wird diese in der längsschnittlichen genetischen Betrachtung nicht weiter berücksichtigt. Tabelle 7.3 fasst die Anzahl an Paaren der längsschnittlichen Betrachtung zusammen.

Tabelle 7.3

Zygotiegruppen der Analysestichproben der ätiologischen Analysen

	ZP	SP	Klasse	Anzahl EZ Paare			Anzahl ZZ Paare			
				m	w	Gesamt	m	w	gg	Gesamt
QS	Z1	K_AS	2 – 4	70	59	129	51	65	116	232
	Z2	K_AS_2	4 – 6	50	44	94	37	53	83	173
	Z3	K_AS_2	6 – 8	50	44	94	37	53	83	173
LS	Z1-Z3	K2_AS_2	3 – 7	18	11	29	18	20	31	69
	Z1-Z3	K3_AS_2	4 – 8	18	16	34	9	20	27	56
Gesamt						63				125

Anmerkungen. EZ = Eineiige Zwillinge; ZZ = Zweieiige Zwillinge; m = männlich; w = weiblich; gg = gegengeschlechtlich.

Für die ätiologischen Analysen lagen entsprechend zwischen 267 und 361 vollständige Paardaten vor (EZ: 94 bis 129; ZZ: 173 bis 232). Die Anzahl an EZ und ZZ Paaren fällt damit insgesamt betrachtet relativ niedrig aus, was sich auf die Power der verhaltensgenetischen Analysen auswirkt und mit großen Konfidenzintervallen der Pfadschätzungen einhergeht (Y. L. L. Luo et al., 2011). Zudem können bei komplexen Modellen instabile Modellschätzungen resultieren. Durch die Zusammenfassung der Kohorten für die längsschnittliche ätiologische Betrachtung wird zudem die Vergleichbarkeit zu den phänotypischen Analysen an einigen Stellen eingeschränkt. Bei der Interpretation der verhaltensgenetischen Ergebnisse ist diesem Umstand entsprechend Rechnung zu tragen.

8 Erhebungsinstrumente

Die Konzipierung des umfangreichen Fragebogensets der ersten Erhebung beruhte teilweise auf Inventaren, die bereits in TEDS eingesetzt wurden. So wurde beispielsweise die Erfassung der kognitiven Fähigkeiten und verschiedener motivationaler Faktoren an die britische Vorgängerstudie angelehnt. Um ein möglichst vollständiges Bild von familiären sowie individuellen Einflussfaktoren auf schulischen Erfolg abzubilden, wurden in KoSMoS zusätzlich weitere Variablen, z. B. das Involvement der Eltern, erfasst (F. M. Spinath & Wolf, 2006).

Im Folgenden wird zunächst das Vorgehen zur Bestimmung der Zygote in KoSMoS geschildert. Anschließend wird die Operationalisierung des schulischen Erfolgs, der allgemeinen kognitiven Fähigkeiten sowie der motivationalen Variablen FSK und IM beschrieben. Für eine ausführliche Beschreibung der ersten Erhebungswelle sei auf Hellbrück (2007) und Toussaint (2007) verwiesen. Das Vorgehen hinsichtlich Anonymisierung und Codierung bei der Online Testung, die den Familien ab dem zweiten Messzeitpunkt zur Wahl stand, findet sich bei Hegewald (2009).

8.1 Zygote

Die Bestimmung der Zygote wurde über einen Elternfragebogen vorgenommen. Obwohl diese Methode hinsichtlich ihrer Klassifikationsgenauigkeit hinter molekulargenetischen Untersuchung mittels DNA-Markern zurücksteht, liefert sie eine zeit- und kostensparende Alternative, vor allem beim Einsatz in großen Probandenstichproben. Der in KoSMoS eingesetzte Fragebogen zur Bestimmung der Zygote stellt eine Adaptation des bereits in TEDS erfolgreich eingesetzten Verfahrens von Goldsmith (1991) dar. Die Eltern beantworteten Fragen zur äußeren Ähnlichkeit der Zwillinge (z. B. hinsichtlich Haar- und Augenfarbe). Darüber hinaus sollte die Verwechslungshäufigkeit im frühen Kindesalter sowie die Veränderung der Ähnlichkeit über das Alter hinweg angegeben werden. Der hieraus errechnete Kennwert der physischen Ähnlichkeit (PSQ; Physical Similarity Quotient) kann Werte zwischen 0 und 1 annehmen, wobei 0 eine maximale physische Ähnlichkeit indiziert. Anhand ermittelter Cut-Off Werte können die Zwillinge dann den Kategorien *eineiig*, *zweieiig* und *nicht feststellbar* zugeordnet

werden. Bisher vorliegende Daten zur Validität des Verfahrens legitimieren dessen Verwendung: Im Vergleich zu einer DNA-basierten Zygotiebestimmung konnte in einer Stichprobe von 18 Monate alten Zwillingen die Eiigkeit mit einer Zuverlässigkeit von 95% festgestellt werden. Auch in einem Retest nach weiteren 18 Monaten erwies sich der Fragebogen als valides Messinstrument: 96% der Paare wurden zu beiden Messzeitpunkten derselben Zygotiegruppe zugeordnet (Price et al., 2000b). Der adaptierte Fragebogen wurde für KoSMoS ins Deutsche übersetzt (siehe Anhang C). Die Verteilung der Zygotie entsprach in allen Kohorten in etwa der Verteilung in der Population (1/3 EZ, 1/3 gleichgeschlechtliche ZZ, 1/3 gegengeschlechtliche ZZ; vgl. Tabelle 7.1), was als Hinweis darauf angesehen werden kann, dass keine selektive Stichprobenauswahl hinsichtlich einer Überrepräsentation von EZ vorlag.

8.2 Erfassung des Kriteriums Schulnoten

Die Schulnoten wurden als übergeordnetes Maß für den schulischen Erfolg und somit auch als Indikator der Leistungsentwicklung herangezogen. Um die Schulnoten der Kinder möglichst vollständig für den gesamten Erhebungszeitraum (von 2005 bis 2009) nutzen zu können, wurden die Eltern zu jedem Erhebungszeitpunkt gebeten, die Noten vollständig für das aktuelle, sowie der zurückliegenden Schuljahre anzugeben. Dies war entweder über ein Onlineformular, telefonisch oder per Paper Pencil Bogen möglich. Insgesamt lagen dementsprechend die Schulnoten der Kinder seit dem Halbjahreszeugnis 2004/05 bis zum Halbjahreszeugnis 2009/10 vor. Die Eltern sollten für jedes Kind (Trennung nach „älterem“ und „jüngeren“ Zwilling) jeweils die Zensuren für die Schulfächer Mathematik, Deutsch (mit der Option die in der Grundschule übliche Unterteilung in Rechtschreibung, Aufsatz und Lesen vorzunehmen), Sachkunde (Grundschule) und die in den Klassenstufen der weiterführenden Schulen gewöhnlich vergebenen Noten für Erdkunde, Biologie und Geschichte angeben.¹ Das Antwortformat folgte der im deutschen Schulsystem üblichen Einteilung der Ziffernoten von eins bis sechs und wurde ergänzt um die Option „nicht erteilt“. Zusätzlich sollten die Eltern angeben, ob sie die Noten dabei aus dem Zeugnis übertragen oder geschätzt haben. Ferner sollte für jedes Schuljahr der von den Kindern besuchte Schultyp mitgeteilt werden.

¹ Die im Folgenden vorgestellten Ergebnisse beziehen sich lediglich auf die Schulnoten der Fächer Mathematik und Deutsch.

8.3 Kognitive Fähigkeiten

Die kognitiven Fähigkeiten der Kinder wurden aufgrund der zeitlich sehr aufwändigen Testung lediglich zum ersten und dritten Messzeitpunkt ermittelt. Durch die geografisch weit verstreute Stichprobe konnte aus Zeit- und Kostengründen keine Face-to-Face Testung der Kinder realisiert werden. Es musste dementsprechend zunächst eine ökonomische und dennoch reliable und valide Erhebungsmethode zur Erfassung der kognitiven Fähigkeiten gefunden werden, um aussagekräftige Ergebnisse zu erhalten. Um dies zu erreichen, wurde in der ersten Erhebung 2005 zunächst auf eine elternadministrierten Heimtestung der Kinder zurückgegriffen, die sich in TEDS bereits bewährt hatte (Toussaint, 2007). Allerdings blieb diese in der deutschen Stichprobe hinter den Erwartungen hinsichtlich ihrer prädiktiven Validität zurück. Daher fand bei der dritten Erhebung 2009 eine Umstellung auf eine telefonbasierte Testung mit geschulten Versuchsleitern statt.

Im Folgenden wird zunächst die Umsetzung beider Erhebungsmethoden genauer beschrieben und hinsichtlich ihrer Güte bewertet. Anschließend werden die eingesetzten Tests erläutert.

8.3.1 Vergleich von Heim- und Telefontestung kognitiver Fähigkeiten

Elternadministrierte Heimtestung

Für die elternadministrierte Testung kognitiver Fähigkeiten in der ersten Erhebung wurden die Eltern ausführlich in einem Informationsschreiben (siehe Anhang C) instruiert und fungierten dann als Versuchsleiter für ihre Kinder. Dieses Vorgehen wurde in KoSMoS aus zweierlei Gründen realisiert: Zum einen ist diese Methode sehr ökonomisch und nahezu kostenneutral. Zum anderen lagen bereits gute Erfahrungen mit dieser Art der Testung in Bezug auf die Reliabilität des Verfahrens aus TEDS vor. Hier muss natürlich mit einer eingeschränkten Durchführungsobjektivität gerechnet werden, die sich wiederum mindernd auf die Reliabilität und die Validität der Messung auswirkt. Eine verminderte Durchführungsobjektivität entsteht in erster Linie durch mögliches Hilfeverhalten der Eltern, sie ergibt sich aber auch aus einer, trotz ausführlicher Instruktion, möglichen Überforderung der Eltern in der Rolle des Versuchsleiters.

In einer Pilotstudie an mehreren Saarbrücker Grundschulen wurde die Heimtestung kognitiver Fähigkeiten mit einer Gruppentestung in der Schule verglichen, um Aufschluss über die Validität des Verfahrens zu erhalten (Pooch, 2006). Nach Ausschluss von Ausreißern lag die Korrelation zwischen Heim- und Schultestung für einen extra-

hierten g -Faktor mit $r = 0.58$ in einem moderaten Bereich. Auf Subtestebene zeigte sich dabei der niedrigste Zusammenhang beim verbalen Untertest „*Allgemeinwissen*“ ($r = 0.33$). In der Studie zeigte sich, dass die interne Validität in Abhängigkeit des Hilfeverhaltens der Eltern variierte. Es lagen besonders dann Einbußen hinsichtlich der Durchführungsobjektivität und damit letztlich auch der Validität (Lienert, 1989) in solchen Fällen vor, bei denen die Eltern angaben, auch in Situationen zu helfen, in denen das Kind eigentlich keine Hilfe benötigt (z. B. bei den Hausaufgaben). Zur Bestimmung der externen Validität wurden die Zusammenhänge zwischen den Schulnoten und der Testung (Heim- vs. Schultestung) herangezogen. Auch hier zeigte sich eine Überlegenheit der Face-to-Face Testung. Die Zusammenhänge zwischen fachspezifischen Noten und g lag hier in allen Fällen bei über 0.60, für die Heimtestung ergab sich eine durchschnittliche Korrelationen von $r = 0.31$ (Pooch, 2006). Insgesamt muss also für eine elternadministrierte Heimtestung kognitiver Fähigkeiten mit gewissen Einbußen hinsichtlich der Validität gerechnet werden. Dies geht wahrscheinlich vor allem auf Unzulänglichkeiten im Hinblick auf die Durchführungsobjektivität zurück (Pooch, 2006).

Ergebnisse basierend auf Daten des ersten Messzeitpunktes lassen ebenfalls auf eine verminderte externe Validität der eingesetzten elternadministrierten Testung kognitiver Fähigkeiten schließen (z. B. für Mathematik $r = 0.33$, vgl. z. B. , Toussaint, 2007). Diese lagen zwar in einem moderaten Bereich, blieben jedoch hinter den in der Literatur berichteten durchschnittlichen Werten von $r = 0.50$ (Fraser et al., 1987; Gustafsson & Undheim, 1996; Neisser et al., 1996) zurück. Aufgrund dieser Befunde wurde für die dritte Erhebung eine Telefontestung durch geschulte Testleiter vorgenommen.

Telefontestung

Eine weitere Möglichkeit, kognitive Fähigkeiten in einer weit verstreuten Stichprobe zu testen, stellt die telefonbasierte Erhebung durch geschulte Testleiter dar. Hierfür wurden in KoSMoS separat verschlossene Umschläge, die das Testmaterial enthielten, in die Familien geschickt, sobald ein Termin für die Testung feststand. Diese Umschläge wurden erst unmittelbar bevor die Testung startete, von den Kindern geöffnet.²

In verschiedenen Studien konnte die Reliabilität und Validität dieses Verfahrens belegt werden. Im Rahmen von TEDS konnten Petrill, Rempell, Oliver und Plomin (2002) in einer Substichprobe von 52 Kindern im Alter von 6 bis 8 Jahren einen substantiellen Zusammenhang zwischen einer Face-to-Face Testung und einer Telefontestung

² Um zu vermeiden, dass die Umschläge bereits vorab geöffnet wurden, erhielten die Eltern die Instruktion, die Umschläge zu verwahren und sie erst dann durch die Kinder öffnen zu lassen, wenn der Testleiter bereits am Telefon war.

von $r = 0.65$ (minderungskorrigiert: $r = 0.75$) nachweisen. Eine in Saarbrücken durchgeführte Studie zur Güte von telefonbasierten Testungen kognitiver Fähigkeiten bestätigte ebenfalls die Reliabilität und Validität eines solchen Verfahrens (Marthaler, 2007). In einer Stichprobe von $N = 98$ Grundschulern im Alter zwischen 7 und 10 Jahren lag der Zusammenhang zwischen Telefontest und einem Schulgruppentest bei $r = 0.81$ und übertraf damit deutlich die bei Pooch (2006) berichteten Zusammenhänge zwischen elternadministrierter Testung und Schulgruppentest. Die Reliabilität der einzelnen Subtests lag durchgängig in einem moderaten bis guten Bereich (zwischen 0.53 und 0.83; Marthaler, 2007).

Die berichteten Befunde sprechen einerseits für eine Überlegenheit der Telefon-
testung im Vergleich zur elternadministrierten Testung, andererseits ist dieses Verfahren weitaus weniger ökonomisch. Jede Testung wird von einem geschulten Testleiter durchgeführt, wofür relativ viele Ressourcen bereitgestellt werden müssen (Terminvereinbarung, Testung von zwei Kindern nacheinander). Aufgrund der als höher zu erwartenden Validität einer telefonbasierten Testung kognitiver Fähigkeiten wurde dieses Vorgehen in der dritten Erhebung dennoch einer elternadministrierten Testung vorgezogen.

8.3.2 Die kognitive Testbatterie

In beiden Erhebungen wurden, basierend auf den in TEDS eingesetzten Inventaren zur Erfassung der kognitiven Fähigkeiten, zwei adaptierte verbale Subtest („Wortschatz“ und „Allgemeinwissen“) sowie zwei adaptierte nonverbale Subtests („Figurenklassifikation“ und „Figurenanalogien“) aus dem KFT 4–12+R (Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision; Heller & Perleth, 2000) und dem HAWIK-III (Hamburg-Wechsler-Intelligenztest für Kinder III; Tewes, Rossmann & Schallberger, 2000) bzw. dem WISC-III (Wechsler Intelligence Scale for Children, 3rd. ed.; Wechsler, 1992) verwendet. Diese wurden in einem *Quizheft* (siehe Anhang C) für die Kinder zusammengestellt.³

In der ersten Erhebung wurden die Quizhefte zusammen mit den übrigen Fragebögen in die Familien geschickt. Eine ausführliche Instruktion über die Durchführung der Testung sowie der Abbruchkriterien für jeden Test wurde für die Eltern beigelegt, um eine möglichst hohe Durchführungsobjektivität zu gewährleisten (siehe Anhang C). Vor jeder Teilaufgabe wurden die Kinder von einem Elternteil entsprechend instruiert und es wurden jeweils zwei Beispielitems mit dem Kind besprochen. Jede Aufgabe musste in einer bestimmten Zeit bearbeitet werden, nach Ablauf der Zeit sollten die Eltern die

³ Der Test ist als Power-Speed-Test zu verstehen, d. h. die Items variieren in ihrer Schwierigkeit und es wird ein Zeitlimit für jeden Untertest vorgegeben.

Bearbeitung beenden. Um mögliche Testängste bei den Kindern zu reduzieren, wurde den Kindern gesagt, dass das Ziel des Quiz die Beurteilung der Schwierigkeit der Items sei und nicht die Testung ihrer Fähigkeiten.

Zum dritten Messzeitpunkt wurden die Fragebögen für die Familien online bereitgestellt⁴, so dass eine separate Verschickung der Quizhefte in die Familien erfolgte. Diese wurden von den Kindern erst zum Zeitpunkt der Erhebung geöffnet. Ein weiterer Unterschied zur ersten Messung liegt in der altersspezifischen Vorgabe der Items. Während zur ersten Messung für alle Kinder die Tests für 4. Klassen eingesetzt wurden, wurden zum dritten Zeitpunkt die jeweiligen Items der entsprechenden Klassenstufe vorgegeben. Hierzu waren die Quizhefte auf der jeweils linken Seite mit Pfeilen versehen, die das Startkriterium (z. B. „Hier beginnt Klasse 5“) eindeutig markierten. Auf der jeweils rechten Seite markierten die Pfeile das Abbruchkriterien (z. B. „Hier endet Klasse 8“). Die Testleiter stellten zusätzlich durch gezieltes Nachfragen sicher, dass die Kinder mit dem richtigen Item begannen. Die Antworten wurden vom Testleiter in ein Antwortformular übertragen. Für die erste Erfassung kognitiver Fähigkeiten muss demnach bei der Interpretation der Befunde in Rechnung gestellt werden, dass die Items für die jüngeren Kinder (Klassenstufe 2 und 3) eventuell zu schwer waren.

Nachfolgend werden die einzelnen Subtests näher beschrieben. In allen Subtests wurden, mit Ausnahme der Skala *Allgemeinwissen*, in beiden Testungen 25 Items präsentiert. In der Skala *Allgemeinwissen* wurden zum ersten Messzeitpunkt 18 und zum dritten Messzeitpunkt 21 Items dargeboten.

- *Quiz 1, Figurenklassifikation (NV1)*

Die Items von Quiz 1 wurden dem nonverbalen Teil des KFT 4–12+R (Heller & Perleth, 2000) entnommen. Aus einer vorgegebenen Reihe von drei oder vier Figuren musste zunächst die zugrunde liegende Gemeinsamkeit extrahiert und anschließend aus einer weiteren Figurenreihe diejenige herausgesucht werden, die zur vorgegebenen Reihe passt. Vorgeschrieben war eine Zeitbegrenzung von 9 Minuten.

- *Quiz 2, Wortschatz (V1)*

Quiz 2 basierte auf den Items des Verbalteils (Form A) des KFT 4–12+R (Heller & Perleth, 2000). Zu einem vorgegebenen Wort sollte aus einer Reihe von fünf weiteren Wörtern das passende Synonym bzw. der passende Oberbegriff gefunden werden. Hierfür standen 7 Minuten zur Verfügung.

- *Quiz 3, Allgemeinwissen (V2)*

⁴ Für Familien ohne Internetzugang stand auch eine Paper Pencil Version zur Verfügung.

Die Items für den zweiten verbalen Subtest wurden in Anlehnung an den HAWIK-III (Tewes et al., 2000) bzw. dem WISC-III (Wechsler, 1992) konzipiert. Die entsprechenden Fragen wurden dem Subtest „Allgemeines Wissen“ des HAWIK-III (Tewes et al., 2000) entnommen. Die zugehörigen Antwortalternativen wurden entsprechend des WISC-III (Wechsler, 1992) adaptiert (ins Deutsche übersetzt und in ein Multiple Choice Format übertragen). In der ersten Erhebung lag das Zeitlimit für diese Aufgabe bei 5 Minuten, für die zweite Erhebung der kognitiven Fähigkeiten war keine Zeitbegrenzung vorgesehen, wohl aber eine definierte Abbruchregel (Abbruch wenn 5 Aufgaben in Folge nicht richtig gelöst werden konnten).

- *Quiz 4, Figurenanalogien (NV2)*

Ebenfalls auf dem nonverbalen Teil des KFT 4–12+R (Form A; Heller & Perleth, 2000) beruhten die Items für das vierte Quiz. Vorgegeben wurde ein Figurenpaar, das in einer bestimmten Relation (Analogie) zueinander stand. Von einem zweiten Paar fehlte eine Hälfte. Aufgabe war es, aus fünf Antwortalternativen diejenige herauszusuchen, die das zweite Paar komplettiert. Aus den Auswahlfiguren sollte also diejenige herausgesucht werden, die zu der dritten Figur passt wie die zweite zur ersten. Für dieses Quiz betrug die Zeitbegrenzung 8 Minuten.

Für die eingesetzten Subtests des KFT 4–12+R (Heller & Perleth, 2000) werden von den Autoren mittlere bis hohe interne Konsistenzen berichtet ($r = 0.83$ bis $r = 0.93$). Für den in Quiz 3 eingesetzten adaptierten Test *Allgemeinwissen* aus dem HAWIK-III (Tewes et al., 2000) wurde in einer Stichprobe von 143 Viertklässlern eine mittlere Reliabilität von $r = 0.62$ (aufgewertet nach Spearman-Brown) ermittelt (Klingbeil, 2004). Es kann weiterhin davon ausgegangen werden, dass die ausgewählten Subskalen ein ausreichend präzises Screening für allgemeine kognitive Fähigkeiten darstellen. Nach Heller und Perleth (2000) lässt sich insbesondere in der Altersgruppe der neun- bis zehnjährigen faktorenanalytisch ein übergeordneter g -Faktor extrahieren, der je nach Stichprobe zwischen 55% und 68% der Varianz aufklärt. Besonders hohe Ladungen (0.87 und 0.93) zeigen dabei die zwei ausgewählten nonverbalen Subtests. Innerhalb der KoSMoS Stichprobe konnte die Extraktion eines g -Faktors aus den vier Subtests ebenfalls bestätigt werden. Spengler (2009) berichtet auf Grundlage von Ergebnissen einer explorativen Faktorenanalyse einen starken unrotierten ersten Faktor, der 59,6% der Varianz aufklärt.

Insgesamt kann davon ausgegangen werden, dass ein Screening kognitiver Fähigkeiten, wie es in KoSMoS realisiert wurde, eine ausreichende Annäherung an die allgemeine Intelligenz darstellt (Klingbeil, 2004).

8.4 Motivation

Die für die Fragestellungen der vorliegenden Arbeit relevanten motivationalen Variablen FSK und IM wurden entsprechend des in Kapitel 3 beschriebenen Modells nach Eccles und Kollegen (Eccles et al., 1983; Eccles & Wigfield, 2002; Wigfield & Eccles, 2000) erfasst. Die Formulierung der Items orientierte sich dabei stark an den Inhalten des jeweiligen Schulfaches. Insgesamt wurden die Bereiche Mathematik, Deutsch, Sachkunde und Sport erfragt, die nachfolgende Darstellung bezieht sich allerdings lediglich auf die beiden relevanten Schulfächer Mathematik und Deutsch. Die Vorgabe der Items erfolgte zu sämtlichen Messzeitpunkten im Selbstbericht. Zusätzlich sollten die Eltern eine Einschätzung über ihre Kinder abgeben. Diese Fremdb Berichte sind für die vorliegende Arbeit jedoch nicht relevant und werden in der Folge nicht weiter betrachtet.

8.4.1 Fähigkeitsselbstkonzept

Die Items zur Erfassung des FSKs bezogen sich im Fach Mathematik auf die Inhalte Kopfrechnen, Textaufgaben und Multiplizieren/Dividieren, im Fach Deutsch wurden die Inhalte Lesen, Aufsatz schreiben und Rechtschreibung erfasst. Die Kinder sollten auf einer fünfstufigen Antwortskala („sehr gut“ bis „gar nicht gut“) einschätzen, *wie gut* sie in diesen Bereichen sind. Die Struktur einer fachspezifischen Unterscheidung konnte innerhalb der KoSMoS-Studie bereits gezeigt werden (Toussaint, 2007). Alle Items der fachspezifischen Skalen können Anhang C entnommen werden.

8.4.2 Intrinsische Motivation

Die Erfassung der IM erfolgte analog zum FSK, d. h. es wurden ebenfalls drei fachspezifische Items zu den gleichen Inhalten erfasst. Die Kinder sollten hier auf einer fünfstufigen Antwortskala („sehr gut“ bis „gar nicht gut“) angeben, wie sehr sie einen bestimmten Bereich innerhalb eines Schulfachs *mögen*. Auch hier konnte die fachspezifische Unterscheidung bereits belegt werden (Toussaint, 2007), alle Items der fachspezifischen Skalen können Anhang C entnommen werden.

9 Analysen der phänotypischen Betrachtung

Die Beantwortung der unter Kapitel 5 formulierten Fragestellungen wird unter Verwendung von latenten Strukturgleichungsmodellen vorgenommen. Die Strukturgleichungsmodellierung (SEM; Structural Equation Modeling) als multivariates statistisches Auswertungsverfahren findet in den empirischen Sozialwissenschaften immer häufigere Anwendung, denn sie bietet ein umfassendes und dabei methodisch adäquates Mittel zur Untersuchung komplexer Zusammenhangsstrukturen. Hierbei bieten sie den Vorteil, dass zufällige Fehlereinflüsse in der latenten Modellierung berücksichtigt und Korrelationen zwischen Variablen simultan betrachtet werden können (Geiser, 2010).

Der Begriff SEM bezeichnet kein singuläres Analyseverfahren sondern eine Gruppe verwandter statistischer Prozeduren. SEM stellt eine integrative konfirmatorische Methode dar, welche die exploratorische Faktorenanalyse mit der Pfadanalyse verbindet (Hoyle, 1995; Kline, 2011). Ziel der SEM ist es, a-priori formulierte Wirkzusammenhänge in einem linearen Gleichungssystem abzubilden und die Modellparameter so zu schätzen, dass die Passung zu den erhobenen Ausgangsdaten maximiert wird (vgl. Weiber & Mühlhaus, 2010, S.17). Hierbei können Zusammenhänge zwischen manifesten¹ und/oder latenten Variablen² analysiert werden. Bei der Betrachtung latenter Variablen, wie sie in dieser Arbeit vorgenommen wird, kann zwischen dem *Messmodell* einer latenten Variable und dem sogenannten *Strukturmodell* unterschieden werden. Während ersteres, dem faktorenanalytischen Ansatz folgend, spezifiziert, welche Indikatoren zur Identifikation des latenten Konstruktes herangezogen werden³, bildet letzteres die Beziehung zwischen den latenten Variablen ab (Weiber & Mühlhaus, 2010). Diese werden als *endogen* bzw. *exogen* bezeichnet, je nachdem, ob sie eine abhängige oder unabhängige (also verursachende) Größe darstellen (Kline, 2011; zur Beziehung zwischen Mess- und Strukturmodell siehe Anhang A). Zur Visualisierung der vermuteten Zu-

¹ Manifeste Variablen sind auf empirischer Ebene direkt beobachtbar, ihre Ausprägungen können mit Messinstrumenten direkt erfasst werden (vgl. Weiber & Mühlhaus, 2010, S.19).

² Latente Variablen sind auf empirischer Ebene nicht direkt beobachtbar, sie werden daher auch als hypothetische Konstrukte bezeichnet (vgl. Weiber & Mühlhaus, 2010, S.19).

³ In diesen auch als *reflektive Messmodelle* bezeichneten Messmodelle wird angenommen, dass die beobachtete Kovarianz zwischen verschiedenen Indikatoren durch die latente Variable verursacht wird (Geiser, 2010).

sammenhangsstrukturen wird auf *Pfaddiagramme* zurückgegriffen, deren Konstruktion einer allgemein anerkannten Konvention folgt (siehe hierzu Anhang B).

Das wesentliche Element bei der Anwendung von Strukturgleichungsmodellen stellt die Überprüfung der Modellgüte des spezifizierten Modells dar. Diese zielt darauf ab, die angestrebte hohe Passung zwischen empirischer Datenlage und geschätztem Modell statistisch zu prüfen. Hierfür stehen verschiedene Fit-Indices zur Verfügung, die unter 9.2 näher erläutert werden. Generell wird bei der Modellprüfung diejenige Lösung angestrebt, die bei geringster Zahl an zu schätzenden Parametern eine noch akzeptable Passung zwischen empirischer und modellimplizierter Kovarianzmatrix erreicht (Parsimonitätskriterium; Schermelleh-Engel, Moosbrugger & Müller, 2003). Im Zuge dessen werden jeweils Submodelle spezifiziert und deren Passung mit einem vollständigen Modell verglichen. Zeigt sich bei einem sparsameren Modell keine signifikante Verschlechterung des Modell-Fits im Vergleich zum vollständigen Modell, wird das sparsamere Modell gewählt (Kline, 2011).

Zur Durchführung von Strukturgleichungsanalysen stehen verschiedene Statistikpakete zur Verfügung. Hier wird auf das Programm MPlus (Version 6; L. K. Muthén & B. O. Muthén, 1998-2010) zurückgegriffen. Mplus bietet den Vorteil, dass der Abhängigkeit der Stichprobe, wie es bei der Verwendung von Zwillingsdaten der Fall ist⁴, bei Anwendung der Funktion TYPE=COMPLEX durch eine Anpassung der Standardfehler Rechnung getragen werden kann (Johnson et al., 2007; L. K. Muthén & B. O. Muthén, 1998-2010). Es wird ein um die Abhängigkeit der Stichprobe korrigierter χ^2 -Wert ausgegeben, dessen Signifikanzprüfung durch die Anpassung der Standardfehler konservativer ausfällt. Zusätzlich wird ein Korrekturfaktor angegeben, der bei der Prüfung von Submodellen verrechnet wird (siehe dazu Abschnitt 9.2; L. K. Muthén & B. O. Muthén, 1998-2010).

Die Anwendung von Strukturgleichungsanalysen ist an bestimmte Voraussetzungen geknüpft, die nachfolgend erläutert werden. Hiernach wird das Vorgehen bei der Modellschätzung und der Überprüfung der Güte der Schätzung vorgestellt. Dabei werden nur diejenigen Indikatoren thematisiert, die zur Bewertung der in Kapitel 12 geschätzten Modelle herangezogen werden. Einen vollständigen Überblick geben hier beispielsweise Schermelleh-Engel et al. (2003), Kline (2011) oder Weiber und Mühlhaus (2010). Anschließend erfolgt eine Einführung in die Analyse von längsschnittlichen Daten wobei hier die Anwendung im MKSD im Mittelpunkt der Betrachtung stehen wird. Eine umfassende Einführung und kritische Auseinandersetzung mit der Methode der Strukturgleichungsmodellierung findet sich z. B. bei Kline (2011).

⁴ Beide Zwillingskinder sind Teil einer Familie.

9.1 Voraussetzungen der Anwendbarkeit der SEM

Die Anwendung von Strukturgleichungsanalysen ist unmittelbar an bestimmte theoretische wie praktische Voraussetzungen geknüpft, deren Verletzung zu Fehlspezifikationen und letztlich nicht interpretierbaren Ergebnissen führt. Zunächst ist es aus theoretischer Sicht unabdingbar, dass bereits a-priori ein Hypothesensystem, welches die vermuteten Wirkzusammenhänge der Variablen spezifiziert, generiert wird (Weiber & Mühlhaus, 2010). Ist dies nicht der Fall, können so lange verschiedene Modellvarianten und Zusammenhänge quasi blind-analytisch modelliert werden, bis ein akzeptabler Fit erreicht wird. Die theoretische Einbindung der gefunden Effekte kann dann nur noch a-posteriori erfolgen – dieser Umstand stellt gleichzeitig eine häufig geäußerte Kritik an strukturgleichungsanalytischen Methoden dar (siehe hierzu auch Kline, 2011).

Zudem müssen auf Seiten praktischer Voraussetzungen die der Analyse zugrunde gelegten Daten vorab hinsichtlich einiger potenziell vorliegender Probleme geprüft werden. So benötigen die meisten Schätzverfahren zur Modellschätzung positiv definite Kovarianzmatrizen (siehe zu diesem Aspekt Kline, 2011, S. 49-51). Die Mehrzahl an zu prüfenden Voraussetzungen betrifft jedoch bestimmte Verteilungsannahmen der eingesetzten Schätzalgorithmen zur Bestimmung der Parameter – die nachfolgend beschriebenen beziehen sich auf die in dieser Arbeit verwendete *Maximum-Likelihood-Schätzung* (ML-Schätzung).

Zunächst ist das Vorliegen *multivariat normalverteilter* Daten zu prüfen. Ist eine Multinormalverteilung der Daten nicht gegeben, kommt es zu starken Verzerrungen der Parameterschätzungen und der Modellgüte. Dies kann in der Folge zu inhaltlich falschen Schlussfolgerungen führen (B. M. Byrne, 2001). Es ist hier jedoch auch in Rechnung zu stellen, dass bei Verwendung von Ratingskalen die strengen Normalverteilungskriterien in aller Regel verletzt sind. Weiber und Mühlhaus (2010) weisen in diesem Zusammenhang darauf hin, dass erst bei gravierenden Abweichungen von der Normalverteilung besagte Verzerrungen in den Parameterschätzungen auftreten, weshalb es im Rahmen der SEM eher gilt, die Stärke der Verletzung der Normalverteilungsannahme zu bewerten. Sofern es sich um eine noch moderate Abweichung handelt, kann trotzdem eine ML-Schätzung verwendet werden, zumal sich dieser Schätzalgorithmus als äußerst robust im Hinblick auf eine Verletzung der Normalverteilungsannahme erwiesen hat (Bollen, 1989; West, Finch & Curran, 1995). Eine Prüfung auf Multinormalverteilung der Daten erfolgt zweistufig. In einem ersten Schritt werden alle Variablen, meist durch Inspektion von Schiefe und Kurtosis, hinsichtlich einer univariaten Normalverteilung geprüft. Als konservative Grenzen werden hier von einigen Autoren Werte x mit $|x| \leq 1$ angegeben, um von einer Nichtverletzung der Normalverteilung ausgehen zu können (z. B. Lienert, 1989). Liberalere Grenzen mit Werten für die Schiefe x_s mit $|x_s| \leq 2$ und der

Kurtosis x_k mit $|x_k| \leq 7$ werden von West et al. (1995) vorgeschlagen. In einem zweiten Schritt erfolgt dann die eigentliche Prüfung auf multivariate Normalverteilung, bei der das Vorliegen univariat normalverteilter Daten zwar eine notwendige, nicht jedoch hinreichende Bedingung darstellt (Kline, 2011). Ein hier häufig herangezogenes Maß ist der sogenannte *Mardia-Koeffizient*. Dieser ist bei Vorliegen einer Multinormalverteilung ebenfalls normalverteilt und kann auf Signifikanz geprüft werden. Resultieren Werte, die signifikant von Null abweichen, muss die Annahme multivariat normalverteilter Daten abgelehnt werden (Weiber & Mühlhaus, 2010). Zur Signifikanzprüfung wird häufig ein *Critical Ratio* (C.R.) herangezogen, der sich durch Division des ermittelten Wertes durch den Standardfehler ergibt. Bei einer strengen Auslegung deuten C.R.-Werte > 1.96 , bei moderater Auslegung Werte > 2.57 auf eine Verletzung der Multinormalverteilung hin (Weiber & Mühlhaus, 2010).

Als weiteren Aspekt der multivariaten Normalverteilung sind die Daten hinsichtlich ihrer *Linearität* und *Homoskedastizität*⁵ durch Inspektion der bivariaten Scatterplots zu prüfen (siehe hierzu z. B. Kline, 2011).

Häufig führen zudem Ausreißer⁶ zu Verletzungen der Normalverteilungsannahme. *Univariate* Ausreißer liegen vor, wenn innerhalb eines Merkmals einzelne ungewöhnliche Werte auftreten, *multivariate* Ausreißer liegen bei auffälligen Wertekombinationen in mehreren Merkmalen vor (Weiber & Mühlhaus, 2010). Als cut-off Wert zur Identifikation univariater Ausreißer werden häufig Werte von mehr oder weniger als drei Standardabweichungen ober- bzw. unterhalb des Mittelwertes angegeben (Tabachnick & Fidell, 2001). Um zu beurteilen, ob Ausreißer einen starken Einfluss auf den Mittelwert haben, kann zusätzlich das 5% getrimmte Mittel herangezogen werden (Pallant, 2010). Hierbei wird der errechnete Mittelwert mit einem Mittelwert verglichen, der sich ergibt, wenn die oberen und unteren 5% der Fälle entfernt werden. Unterscheiden sich beide Werte nur marginal voneinander, kann davon ausgegangen werden, dass die Extremwerte nur von geringer Bedeutung sind. Die Entscheidung, ob Ausreißer letztlich aus dem Datensatz entfernt werden, sollte jedoch nicht nur auf Grundlage statistischer Kennwerte, sondern immer auch vor dem Hintergrund inhaltlicher Überlegungen getroffen werden (Bühner, 2006). Die Identifikation multivariater Ausreißer erfolgt unter Rückgriff auf die sogenannte *Mahalanobis-Distanz*. Diese prüft für jeden Fall, wie stark er vom multivariaten Mittelwert abweicht. Liegen ausreichend große Stichproben vor, sind die Distanzwerte χ^2 verteilt, wobei die Anzahl an Freiheitsgraden (degrees of freedom; df) der Anzahl an Variablen entspricht. Jeder sich ergebende Distanzwert kann dann im Zuge einer Signifikanzprüfung mit einem kritischen χ^2 -Wert ($p < 0.01$) vergli-

⁵ *Homoskedastizität* bedeutet, dass die Varianz der Residuen einer Variable (bzw. der Variable selbst) für alle Stufen einer anderen Variable gleich ist (Bortz, 1999).

⁶ Ausreißer sind Beobachtungswerte, die aus sachlogischer Sicht als ungewöhnlich, nicht plausibel und widersprüchlich zu bezeichnen sind und die dadurch nicht zu den übrigen Beobachtungswerten einer Variablen bzw. deren Verteilung passen (vgl. Weiber & Mühlhaus, 2010, S.145).

chen und kritische Fälle identifiziert werden (Kline, 2011).

Neben einer Verletzung der Normalverteilungsannahme kann auch das Vorliegen einer extremen *Kollinearität* zu instabilen und nicht interpretierbaren Parameterschätzungen führen. Als Richtwert einer hohen Kollinearität werden quadrierte multiple Korrelationen (R_{mc}^2) von > 0.90 angegeben (vgl. z. B. Tabachnick & Fidell, 2001). Um die Variablen auf Kollinearität zu prüfen, kann neben der Inspektion der bivariaten Korrelationen der *Toleranz-Wert* (TW) sowie der *Variance Inflation Factor* (VIF) betrachtet werden. Der Toleranz-Wert gibt denjenigen Varianzanteil an, der nicht durch die Summe der übrigen Variablen erklärt wird ($TW = 1 - R_{mc}^2$). Werte $TW < 0.10$ zeigen eine extreme Multikollinearität an. Der VIF setzt die gesamte standardisierte Varianz ins Verhältnis zur spezifischen Varianz:

$$VIF = \frac{1}{1 - R_{mc}^2}$$

Werte von > 10 deuten auf eine Redundanz der Variablen hin (Kline, 2011).

Ein weiterer Aspekt den es zu prüfen gilt stellt die *Varianzhomogenität* der Variablen dar. Unterscheiden sich die Varianzen sehr stark, kann dies zum Fehlschlagen des Iterationsprozesses führen (Kline, 2011). Werden unterschiedliche Gruppen miteinander verglichen, sollten weiterhin keine Unterschiede in den Varianzen der Gruppen vorliegen. Die Prüfung dieser Annahme erfolgt durch den Levene-Test. Tabachnick und Fidell (2001) empfehlen bei inhomogenen Varianzen die Durchführung eines F_{max} -Tests, um das Ausmaß der Varianzinhomogenität abschätzen zu können. Hierzu wird die größte Gruppenvarianz ins Verhältnis zur kleinsten Gruppenvarianz gesetzt. Bei relativ gleichen Gruppengrößen (maximal 4:1) darf der F_{max} -Test keinen Wert größer als 10 ergeben um von der Homogenität der Varianzen ausgehen zu können (Tabachnick & Fidell, 2001).

Letztlich ist zu berücksichtigen, dass zur Durchführung von SEM eine gewisse *Stichprobengröße* gewährleistet sein muss, um stabile Kovarianzmatrizen zu erhalten, die als Grundlage zur Modellschätzung (vgl. Abschnitt 9.2) herangezogen werden. Dabei gilt allgemein, dass mit zunehmender Komplexität des zu schätzenden Modells auch die Stichprobe entsprechend groß sein muss (Kline, 2011). Allerdings gibt es hierbei keine feste Regel, ab wann eine Stichprobe als nicht mehr ausreichend groß gilt. Als nützlich hat sich hierbei die von Jackson (2003) vorgeschlagene *N:q Regel* erwiesen, nach der das Verhältnis zwischen Fallzahl und zu schätzenden Parametern im Idealfall 20:1 aber zumindest 10:1 betragen sollte (siehe Kline, 2011). Als Folge zu geringer Stichprobengrößen reduziert sich insgesamt die Power der Analyse (größere Standardfehler), was in numerisch betrachtet bedeutsamen, statistisch jedoch nicht signifikanten Parameterschätzungen resultieren kann (Kline, 2011).

Da die meisten Schätzalgorithmen vollständige Datensets voraussetzen (Hoyle, 1995),

ist die Frage nach dem Umgang mit *fehlenden Werten* bei der Anwendung von Strukturgleichungsanalysen von besonderer Relevanz. In diesem Zusammenhang ist in erster Linie zu prüfen, ob ein systematischer Ausfall der Daten vorliegt. Die absolute Anzahl an fehlenden Werten ist dabei weniger bedeutsam (Weiber & Mühlhaus, 2010).⁷ Rubin (1976) unterscheidet drei Ausfallmuster, die mit Hilfe statistischer Verfahren geprüft werden können. Fallen die Werte *not missing at random* (NMAR) aus, hängt die Wahrscheinlichkeit einer fehlenden Angabe in einer Variable x von deren „wahren“, aber nicht beobachtbaren Wert selbst ab (Weiber & Mühlhaus, 2010). Fallen die Werte vollständig zufällig aus (*Missing completely at random*; MCAR) steht das Fehlen der Werte in keinerlei Zusammenhang zu anderen fehlenden oder vorhandenen Werten (Weiber & Mühlhaus, 2010). Liegt ein *Missing at random* (MAR) Muster vor, bedeutet dies, dass die Ausfallwahrscheinlichkeit einer Variable x zwar von anderen Variablen, nicht aber von der Variable x selbst abhängt (Weiber & Mühlhaus, 2010). Liegen MCAR bzw. MAR Ausfallmuster vor, können die fehlenden Werte unter Verwendung adäquater Schätzmethoden (siehe hierzu im Detail z.B. Allison, 2002) ersetzt werden. Verfahren, die auf ML-Schätzungen basieren, werden dabei mittlerweile gängigen Verfahren, wie z. B. der Ersetzung durch den Mittelwert, vorgezogen. Im Rahmen dieser Arbeit wird hierfür der *Expectation Maximization (EM) Algorithmus* verwendet.⁸ Die Prüfung der Verteilung fehlender Werte wird üblicherweise unter Verwendung des MCAR-Tests nach Little (R. J. A. Little, 1988) vorgenommen. Ergibt dieser nicht signifikante Werte, gilt die MCAR Bedingung als bestätigt. Wird der MCAR-Test signifikant, wird in einem nächsten Schritt die Haltbarkeit der MAR Bedingung durch eine t-Statistik zwischen Itempaarungen geprüft.

An dieser Stelle ist zudem wichtig zu erwähnen, dass es gerade im Kontext von Längsschnittanalysen von großer Bedeutung ist, fehlende Werte adäquat zu ersetzen, um eine systematische Verzerrung der Daten über den Untersuchungsverlauf hinweg zu vermeiden (siehe Allison, 2002, S.74). Allison (2002) konnten in diesem Zusammenhang in verschiedenen Simulationen zeigen, dass bis zu 50% gelöschter Daten durch ML-Schätzungen ersetzt werden können, ohne dass sich dies auf die Ergebnisse auswirkt.

⁷ Kline (2011) gibt hier an, dass eine Anzahl von 5% fehlender Werte in den Daten vernachlässigbar sind.

⁸ Der EM-Algorithmus stellt eine zweischrittige iterative Prozedur dar: Nach der Schätzung (Estimation) fehlender Werte durch eine Abfolge von Regressionen wird eine ML Schätzungen (Maximization) des vervollständigten Datensatzes durchgeführt. Diese Schritte werden so lange wiederholt, bis eine stabile Lösung erreicht wird (Kline, 2011).

9.2 Modellschätzung und Güteprüfung

Wie eingangs bereits beschrieben, wird, nach erfolgreicher Voraussetzungsprüfung, bei der Durchführung einer SEM zunächst ein Hypothesensystem aufgestellt, dass die vermuteten Wirkzusammenhänge zwischen den betrachteten Variablen spezifiziert. Für jede abhängige Variable wird dann genau eine lineare Gleichung formuliert, so dass das gesamte Hypothesensystem in ein lineares Mehrgleichungssystem überführt wird (Kline, 2011). Die eigentliche Modellschätzung erfolgt auf Basis der Informationen aus der empirischen Varianz-Kovarianzmatrix (Weiber & Mühlhaus, 2010). Damit die Parameter geschätzt werden können, muss das Modell mindestens so viele bekannte wie zu schätzende Parameter enthalten. In diesem Fall liegt ein *gerade identifiziertes Modell* mit Null Freiheitsgraden vor, wodurch die Prüfung der Modellpassung nicht möglich ist. Liegen weniger bekannte als zu schätzende Parameter vor, können in diesem *unteridentifizierten Modell* auch keine Parameterschätzungen vorgenommen werden. Hingegen kann in einem *überidentifizierten Modell* mit mehr bekannten als zu schätzenden Parametern ($df \geq 1$) neben der Schätzung der Parameter auch die Passung der modellimplizierten zur empirischen Varianz-Kovarianzmatrix geprüft werden (Hoyle, 1995). Dabei wird eine minimale Diskrepanz beider angestrebt (Kline, 2011). Das am häufigsten eingesetzte Schätzverfahren stellt hier der ML-Algorithmus dar, mit welchem in einem iterativen Prozess die Diskrepanzfunktion minimiert wird (für eine Erläuterung der ML-Schätzung sei an dieser Stelle auf Reinecke (2005) verwiesen). Wie gut diese Annäherung an die empirischen Daten gelingt, kann durch unterschiedliche Fit-Indikatoren bestimmt werden. Der Beurteilung des Modell-Fits geht im Allgemeinen die Prüfung der Plausibilität der Schätzungen und der Größe der Standardfehler voraus (Hoyle, 1995).

Die Frage, welche spezifischen Indikatoren zur Evaluation des Modells und potenzieller Submodelle aus der Vielzahl an zur Verfügung stehenden Kennwerten herangezogen werden sollten, lässt sich nicht eindeutig im Sinne einer allgemeingültigen Regel beantworten. Kline (2011) diskutiert diesen Aspekt der Abhängigkeit der Interpretation von SEM Ergebnissen von Stichprobe und ausgewählter Fit-Statistik sowie Vor- und Nachteile bestimmter Indikatoren kritisch (vgl. S. 189ff). In der Literatur finden sich daneben zahlreiche Empfehlungen, welche Kriterien zur Beurteilung der Modellgüte herangezogen werden sollten. Grundsätzlich wird jedoch angeraten, sich nicht auf die Angabe eines einzelnen Indikators zu beschränken, sondern eine Zusammenstellung verschiedener Kennwerte heranzuziehen (Schermelleh-Engel et al., 2003). Weiber und Mühlhaus (2010) fassen die Kennzeichen eines *guten* Modells wie folgt zusammen (vgl. S. 176): Die empirische Varianz-Kovarianzmatrix sollte möglichst fehlerfrei vorhergesagt werden können. Diese Forderung kann durch *absolute Fitindizes* geprüft werden.

Das Modell sollte insgesamt mit möglichst wenigen zu schätzenden Parametern auskommen, also dem Parsimonitätskriterium entsprechen und schließlich erheblich besser fitten als ein Modell, welches die Beziehungen im Datensatz als zufallsbedingt ansieht (*Basismodell*). Für letzteres können *inkrementelle Fitmaße* herangezogen werden. Hu und Bentler (1999) sprechen sich auf Grundlage der Ergebnisse einer durchgeführten Simulationsstudie für die gemeinsame Anwendung der Indikatoren *TLI* (Tucker-Lewis-Index), *RNI* (Relative Noncentrality Index) oder *CFI* (Comparative Fit Index) in Verbindung mit dem *SRMR* (Standardized Root Mean Square Residual) aus, da sich hier in der Summe die geringsten Fehler 1. und 2. Art gezeigt haben (zitiert nach Weiber & Mühlhaus, 2010). Zusätzlich sollte der χ^2 -Wert des Modells betrachtet werden, da dieser das einzige inferenzstatistische Maß darstellt, welches den Fit des Gesamtmodells anhand eines Signifikanztests prüft (Schermelleh-Engel et al., 2003). In verhaltensgenetischen Analysen wird im Allgemeinen der *RMSEA* (Root-Mean-Square-Error of Approximation) betrachtet. Die Güteprüfung der unter Kapitel 14 geschätzten Modelle folgt den Empfehlungen von Hu und Bentler (1999) sowie Schermelleh-Engel et al. (2003). Die verwendeten Fitmaße werden im Folgenden einer näheren Betrachtung unterzogen.

Mit dem χ^2 -Test wird die Nullhypothese, dass empirische und modellimplizierte Kovarianzmatrix identisch sind, geprüft (Weiber & Mühlhaus, 2010). Die χ^2 -Werte fallen umso höher aus, je größer die Diskrepanz beider Matrizen ist. Jedem χ^2 -Wert kann eine Wahrscheinlichkeit p zugeordnet werden, mit der die Ablehnung der Nullhypothese eine Fehlentscheidung darstellt (Weiber & Mühlhaus, 2010). Mit dem χ^2 -Wert sind jedoch einige Schwierigkeiten verbunden. Zum einen ist er von der Stichprobengröße in der Art abhängig, dass bei wachsender Stichprobengröße und gleichbleibender Zahl an Freiheitsgraden der χ^2 -Wert ansteigt und signifikant wird – wodurch ein eigentlich passendes Modell abgelehnt werden würde (Kline, 2011). Im umgekehrten Fall kleiner Stichproben sinken die χ^2 -Werte (und der damit verbundene p -Wert) was zur Folge haben kann, dass auch Modelle mit großer Diskrepanz zwischen beobachteter und modellimplizierter Matrix als passend angenommen werden (Schermelleh-Engel et al., 2003). Eine weitere Schwierigkeit ist darin zu sehen, dass der χ^2 -Wert komplexe Modelle bevorzugt (je mehr Parameter, desto weniger Freiheitsgrade, desto geringer der χ^2 -Wert) was dem Parsimonitätskriterium widerspricht (Schermelleh-Engel et al., 2003). Es wird daher empfohlen, den χ^2 -Wert lediglich als deskriptives Maß zu interpretieren und ihn nicht im Sinne einer Inferenzstatistik zu nutzen. Zudem besteht die Möglichkeit, den χ^2 -Wert an der Anzahl an Freiheitsgraden zu relativieren (χ^2/df ; siehe dazu z. B. Kline, 2011). Bleibt dieses Verhältnis unter einem Wert von 2, liegt ein guter Fit, bei Werten < 3 ein noch akzeptabler Fit vor (Schermelleh-Engel et al., 2003). Um die Nachteile der χ^2 -Statistik zu umgehen, wurden verschiedene Fit-Indikatoren entwickelt. Diese leiten sich zumeist aus dem χ^2 -Wert ab und bewerten entweder den

Gesamtfit des Modells, die Güte eines Modells in Relation zu Alternativmodellen oder die Sparsamkeit des Modells (Schermelleh-Engel et al., 2003).

Der TLI (Bollen, 1989) berücksichtigt neben der Differenz der χ^2 -Werte des formulierten Modells M und des Basismodells B die Freiheitsgrade beider Modelle. Die Gültigkeit der χ^2 -Verteilung wird dabei vorausgesetzt.

$$\text{TLI} = \frac{\frac{\chi_B^2}{df_B} - \frac{\chi_M^2}{df_M}}{\frac{\chi_B^2}{df_B} - 1}$$

χ_M^2 = Chi-Quadrat des formulierten Modells

χ_B^2 = Chi-Quadrat des Basismodells

df_M = Freiheitsgrade des formulierten Modells

df_B = Freiheitsgrade des Basismodells

Nimmt der TLI Werte > 1 an, deutet dies darauf hin, dass im formulierten Modell mehr Parameter als notwendig spezifiziert wurden (Weiber & Mühlhaus, 2010). Als Schwellenwert für einen guten Modell-Fit gelten Werte ≥ 0.90 .

Der CFI (Bentler, 1990) testet die Modellgüte im Vergleich zum *independence model*⁹ und bewertet die Verbesserung der Anpassung an die empirischen Daten durch das theoretische Modell. Der CFI ist auf das Intervall $[0; 1]$ normiert, Werte ≥ 0.95 indizieren einen akzeptablen, Werte ≥ 0.97 einen guten Modell-Fit (Schermelleh-Engel et al., 2003).

$$\text{CFI} = 1 - \frac{\chi_M^2 - df_M}{\chi_B^2 - df_B}$$

χ_M^2 = Chi-Quadrat des formulierten Modells

χ_B^2 = Chi-Quadrat des Basismodells

df_M = Freiheitsgrade des formulierten Modells

df_B = Freiheitsgrade des Basismodells

Der SRMR stellt ein absolutes Fitmaß dar, welches den Differenzwert zwischen empirischer und modellimplizierter Varianz-Kovarianzmatrix in Relation zur Komplexität eines Modells setzt. Je kleiner die SRMR-Werte, desto besser gelingt die Anpassung

⁹ Alle Parameter dieses Modells mit Ausnahme der Varianzen der manifesten Variablen sind auf Null fixiert.

an die empirischen Daten. Als Grenzwert gilt ein SRMR-Wert ≤ 0.10 als akzeptabler, Werte ≤ 0.08 als guter Modell-Fit (Hu & Bentler, 1999).

$$\text{SRMR} = \sqrt{\frac{2 \sum \sum \left(\frac{s_{ij} - \sigma_{ij}}{s_{ii} s_{jj}} \right)^2}{p(p+1)}}$$

s_{ij} = Empirische Varianz-Kovarianz der Variablen x_{ij}

σ_{ij} = Modelltheoretisch errechnete Varianz-Kovarianz
der Variablen x_{ij}

p = Anzahl der Indikatoren

Der RMSEA ist ein deskriptiver Index der relativ unabhängig von der Stichprobengröße ist. Der RMSEA berücksichtigt die Komplexität eines Modells und belohnt parsimonischere Modellierungen (Weiber & Mühlhaus, 2010). Ein Wert ≤ 0.08 kennzeichnet einen mittleren, ein Wert ≤ 0.05 einen guten und ein Wert ≤ 0.01 einen sehr guten Fit (Neale et al., 2005). Zusätzlich kann für den RMSEA ein Intervall berechnet werden, welches angibt, wie hoch die Wahrscheinlichkeit ist, dass der RMSEA < 0.05 ist (Kline, 2011). Der RMSEA wird in vorliegender Arbeit zur Bewertung des Fits der verhaltensgenetischen Modelle (vgl. Kapitel 10) sowie der Modelle im MKSD herangezogen, da der SRMR hier bei kleinen Stichproben und großer Zahl an Freiheitsgraden verzerrte Schätzungen liefern würde (Kenny, 2011). Der RMSEA berechnet sich als

$$\text{RMSEA} = \sqrt{\frac{\chi_M^2 - df_M}{df_M(n-1)}}$$

χ_M^2 = Chi-Quadrat des formulierten Modells

df_M = Freiheitsgrade des formulierten Modells

n = Anzahl der Probanden

Um der Forderung sparsamer Modelle nachzukommen, können durch Fixierung bestimmter Parameter Submodelle des theoretischen Modells getestet werden. Es wird dabei generell geprüft, ob ein solches *genestetes Modell*¹⁰ die Daten ohne signifikante Verschlechterung des Modell-Fits angemessen beschreiben kann. Ein Vergleich genesteter Modelle kann mittels des χ^2 -Differenztests (Likelihood-Ratio-Test; LRT), welcher die Differenz im Modell-Fit (bezogen auf den χ^2 -Wert mit $df = \Delta df$) auf Signifikanz

¹⁰ Zwei Modelle sind genau 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).

prüft, erfolgen. Da auch die Differenz zweier χ^2 -Werte einer χ^2 -Verteilung folgt, kann die Nullhypothese, dass beide Modelle eine gleich gute Passung aufweisen, geprüft werden. Fittet das genestete Modell nicht signifikant schlechter ($p \geq 0.05$) wird das sparsamere Modell dem vollständigen Modell vorgezogen (Neale & Cardon, 1992).

$$\text{LRT}(\Delta df) = \chi_A^2(df_A) - \chi_B^2(df_B)$$

$$\Delta df = df_A - df_B$$

B = Vollständiges Modell

A = Genestetes Modell

Ein Sonderfall ergibt sich bei Vorliegen geclusterter Daten. Hierbei wird zur vergleichenden Prüfung genesteter Modelle der von Mplus ausgegebene Korrekturfaktor benötigt. Um zur benötigten Prüfstatistik zu gelangen, wird in diesem Fall der *Satorra-Bentler Chi-Quadrat Differenztest* (TRd; Satorra & Bentler, 2001) anstatt des LRTs berechnet.¹¹

$$\text{TRd} = \frac{T_0 \times c_0 - T_1 \times c_1}{cd}$$

wobei gilt

$$cd = \frac{df_0 \times c_0 - df_1 \times c_1}{df_0 - df_1}$$

T_0 = Chi-Quadrat-Wert des genesteten Modells

T_1 = Chi-Quadrat-Wert des Vergleichsmodells

c_0 = Korrekturfaktor des genesteten Modells

c_1 = Korrekturfaktor des Vergleichsmodells

df_0 = Freiheitsgrade des genesteten Modells

df_1 = Freiheitsgrade des Vergleichsmodells

Sollen zwei Modellalternativen verglichen werden die nicht genestet sind, werden andere Kennwerte herangezogen. Das *Akaike Information Criterion* (AIC; Akaike, 1987) berücksichtigt den χ^2 -Wert und die Anzahl an Freiheitsgraden des Modells. Dabei wird dasjenige Modell mit dem niedrigeren AIC Wert bevorzugt (Schermelleh-Engel et al., 2003).

¹¹ In diesem Fall wird anstelle einer ML-Schätzung eine robuste ML-Schätzung (MLM) in Mplus verwendet (Christ & Schlüter, 2012; L. K. Muthén & B. O. Muthén, 1998-2010).

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

df = Freiheitsgrade des Modells

Resultiert ein nur mäßiger Modell-Fit können unter Zuhilfenahme sogenannter *Modification Indices* sukzessiv Parameterschätzungen zugelassen werden, die zu einer Verbesserung des Fits führen. Hierauf wird an dieser Stelle nicht weiter eingegangen, ausführlich beschrieben ist dies z. B. bei Bühner (2006). Außerdem sei noch darauf hingewiesen, dass ein guter Fit alleine noch nicht ausreichend ist, um von der Angemessenheit eines Modells zur Beschreibung beobachteter Daten ausgehen zu dürfen. Ein guter Modell-Fit kann prinzipiell auch durch die Berücksichtigung weiterer Parameter erreicht werden (Reduzierung der Freiheitsgrade), wobei ein solches Modell keinen Wert im wissenschaftlichen Sinne mehr hat (Kline, 2011).

Zur Überprüfung der unter 5.2 formulierten Fragestellung wird eine CLP-Analyse durchgeführt, deren Annahmen und Voraussetzungen nachfolgend erläutert werden. Daneben wurden auch zur Überprüfung eines REM auf Grundlage einer CLP-Analyse bestimmte Richtlinien erarbeitet (Marsh et al., 1999). Die Besonderheiten einer Ausweitung auf ein MKSD werden nach dieser allgemeinen Einführung näher ausgeführt. Abschließend wird die Überprüfung der Messinvarianz als notwendige Voraussetzung längsschnittlicher SEM Analysen bzw. Gruppenvergleichen beschrieben.

9.3 SEM im Rahmen längsschnittlicher Analysen

Gerade bei der Analyse von Veränderungsprozessen im Kindes- und Jugendalter ist in den letzten Jahrzehnten eine Verschiebung von Querschnitts- hin zu Längsschnittanalysen zu beobachten, da nur diese die Beantwortung der Frage nach Stabilität und Veränderung in bestimmten Merkmalen erlauben (Reinders, 2006; Schmiedek & Lindenberger, 2007). Längsschnittdaten sind hier in besonderer Weise geeignet, Entwicklungsprozesse von Kohorteneffekten entkoppelt zu betrachten. Nach Reinders (2006) können empirisch gesehen zum einen Fragen nach dem *Was* und *Wie* von Veränderung durch einen Zeitvergleich von Mittelwerten und Korrelationen beantwortet werden. Zum anderen stellt sich auch die Frage nach dem *Warum* der Veränderung. Letzteres bezieht sich auf die Kausalität zwischen unabhängigen (UV) und abhängigen Variablen (AV), sprich der *Ursachen* von Entwicklungsprozessen. Ein Nachweis der Kausalität ist dann gegeben, wenn die Ursache der Wirkung zeitlich vorgelagert ist, beide kovariieren

und die Ursache die hauptsächliche Erklärung der Wirkung darstellt (vgl. Reinders, 2006). Die Wirkung stellt hier die relative Veränderung der AV über die Zeit hinweg, und nicht etwa deren absolute Ausprägung zu Zeitpunkt 2, dar (Schmiedek & Lindenberger, 2007). Zum empirischen Nachweis von Kausalität muss dementsprechend die Kovariation von UV und veränderten Ausprägungen der AV geprüft werden (Reinders, 2006). Zur Identifikation von Kausalstrukturen wurde das CLP-Design entwickelt (Lazarsfeld, 1972; zitiert nach Kenny, 1975). Hierbei hat sich das methodische Prinzip seither nicht verändert, zur Auswertung von Daten im CLP-Design stehen allerdings mittlerweile durch die Weiterentwicklung strukturgleichungsanalytischer Verfahren adäquatere Methoden zur Verfügung.

Das Grundprinzip der CLP-Analyse ist in Abbildung 9.1 skizziert. Es handelt sich demnach um ein Modell, welches die Beziehung zwischen zwei Variablen zu zwei Zeitpunkten¹² expliziert.

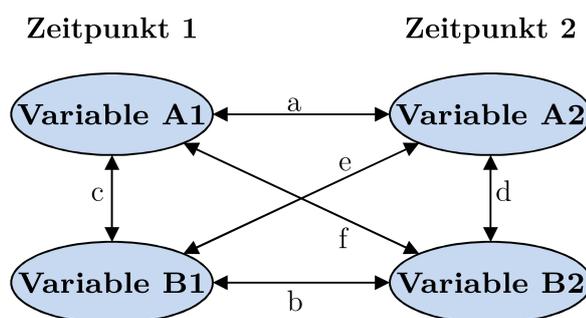


Abbildung 9.1. Schematische Darstellung eines cross-lagged panel Designs mit zwei Variablen zu zwei Zeitpunkten. Eigene Darstellung in Anlehnung an Abbildung 1 aus Reinders, 2006, S. 572.

Es können insgesamt drei Korrelationen unterschieden werden (siehe dazu z. B. Kenny, 1975; Reinders, 2006):

- Die *Autokorrelationen* der Variablen A und B über die Zeit (Pfade a und b) repräsentieren die mittlere Stabilität der Variablen über die Zeit.
- Die *zeitgleichen Korrelationen* (Pfade c und d) stellen die Höhe des mittleren linearen Zusammenhangs zwischen den Variablen A und B innerhalb eines Zeitpunktes dar.
- Die *Kreuzkorrelationen* (engl. *cross-lagged correlation*; Pfade e und f) und deren Relation zueinander dienen schließlich der Überprüfung der Kausalität der Varia-

¹² Eine Erweiterung auf mehr als zwei Variablen oder mehr als zwei Zeitpunkte ist möglich, stellt aber höhere Anforderungen an die dahinter liegende Theorie und die empirische Datengrundlage (Reinders, 2006).

blen A und B. Fällt einer der beiden Pfade substanziell höher aus als der andere, wird von einem kausalen Zusammenhang in entsprechender Richtung ausgegangen.

Die postulierten korrelativen Zusammenhänge sollten dabei der Forderung nach *Gleichzeitigkeit* der Messung (engl. *synchronicity*) sowie der *Stationarität* (engl. *stationarity*) genügen (Kenny, 1975). Stationarität bezieht sich auf die Gleichheit der Konstrukte zu unterschiedlichen Zeitpunkten im Sinne der Konstruktoperationalisierung (zu $Z1$ wird nichts anderes erfasst als zu $Z2$) sowie der Gleichheit von Stärke und Richtung der Ursachen dieser Variable (Kenny, 1975). Hiervon zu unterscheiden ist die *Stabilität* der Variablen (Pfade a und b in Abbildung 9.1).

Das CLP Modell war ursprünglich darauf ausgelegt, die Zusammenhänge mittels einfacher bivariater Korrelationen zu prüfen, wodurch die Frage nach *Veränderung* der Variablen über die Zeit nicht beantwortet werden kann. Dem Wirkungsbegriff kann auf diese Weise demnach keine Rechnung getragen werden (vgl. z. B. Rothermund, 2003). Erreicht werden kann dies durch Kontrolle des Einflusses der Autokorrelation, wodurch die verbleibende Varianz der Variable A_2 (kontrolliert um den Einfluss von A_1) als Veränderungsvarianz plus Messfehler interpretiert werden kann (Reinders, 2006). Es wird entsprechend die Höhe der Korrelation von Variable B_1 mit dieser Restvarianz betrachtet. Findet sich ein Zusammenhang, ist die Veränderung von A über die Zeit auf die Ausprägung der Variable B_1 zurückführbar. Der Einfluss von Variable B_1 hängt folglich von der Höhe der Autokorrelation von Variable A ab: Je höher diese ausfällt, desto geringer der potenzielle Einfluss von Variable B (Reinders, 2006). Ein solches Modell ist in Abbildung 9.2 dargestellt.

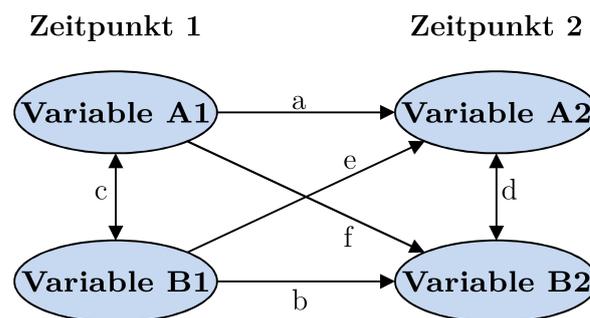


Abbildung 9.2. Cross-lagged panel Design mit Kontrolle der Autokorrelation. Eigene Darstellung in Anlehnung an Abbildung 3 aus Reinders, 2006, S. 574.

Im CLP-Design kann also nicht nur eine Alternativhypothese (A führt zu B) getestet werden, es ist vielmehr möglich, zwei miteinander konkurrierende (sich aber nicht notwendigerweise gegenseitig ausschließende) Modelle simultan zu prüfen (Schmiedek & Lindenberger, 2007), was die besondere Eignung dieses Designs zur Überprüfung des

REMs herausstellt.

Das bereits in 3.4.1 eingeführte prototypische Modell zur Überprüfung des REM (vgl. Abbildung 3.1) basiert auf den zuvor beschriebenen Grundlagen des CLP-Designs. Marsh et al. (1999) empfehlen allgemeine Richtlinien, die bei der Durchführung eines REMs zur Überprüfung der *skill-development* vs. *self-enhancement* Hypothese beachtet werden sollten:

- Die Operationalisierung von FSK und SE sollte auf Basis von mindestens drei Indikatoren erfolgen.
- Die Kontrolle von Methodeneffekten über die Zeit (*correlated uniqueness*) ist bei Verwendung der gleichen Instrumente zu unterschiedlichen Zeitpunkten zwingend erforderlich, um eine Überschätzung der Stabilität der Variablen zu vermeiden (Marsh & Hau, 1996).
- FSK und SE sollten zu mindestens zwei, besser jedoch zu mehr als zwei Zeitpunkten, erhoben werden. Der zeitliche Abstand sollte dabei konstant gehalten werden.
- Die Modellprüfung sollte immer mit einem vollständigen Modell (engl. *Full-Forward Model*) beginnen, in dem sämtliche Autokorrelationen, zeitgleichen Korrelationen und Kreuzkorrelationen geschätzt werden. Alle weiteren zu testenden Modelle sind in diesem genestet, was die Signifikanzprüfung unterschiedlicher Submodelle ermöglicht.
- Um eine stabile Modellschätzung zu erhalten, sollten die untersuchten Stichproben eine Mindestgröße aufweisen. Hierzu werden von Marsh et al. (1999) keine weiteren Angaben gemacht, es wird dementsprechend den Empfehlungen von Jackson (2003) gefolgt (siehe dazu Abschnitt 9.1).

Die Prüfung beider Modelle erfolgt dann durch sukzessive Fixierung bestimmter Pfade und den Vergleich der resultierenden, restringierten Modelle gegen das vollständige Modell. Auf diese Weise kann die Signifikanz der Kreuzpfade unter Kontrolle der Autokorrelationen und der zeitgleichen Korrelationen geprüft werden. Die Abfolge der zu schätzenden Modelle wird in den Ergebnissen der phänotypischen Analysen beschrieben (vgl. Kapitel 13).

Innerhalb dieser Arbeit findet eine Ausweitung der Prüfung des REMs (neben einer Übertragung auf IM und der fachspezifischen Betrachtung) auf ein MKSD statt. Abbildung 9.3 stellt ein prototypisches MKSD für drei Kohorten und zwei (durchgezogene Linien) bzw. drei (gestrichelte Linien) Messzeitpunkte dar. Durch die Berücksichtigung

mehrerer Alterskohorten innerhalb einer längsschnittlichen Analyse kann der Einfluss von Kohorten- und Entwicklungseffekten abgeschätzt werden. Hierfür werden Kohorten mit mindestens zwei überlappenden Zeitpunkten benötigt (in KoSMoS sind dies die Kohorten 1 und 3 in den Klassenstufen vier und sechs).

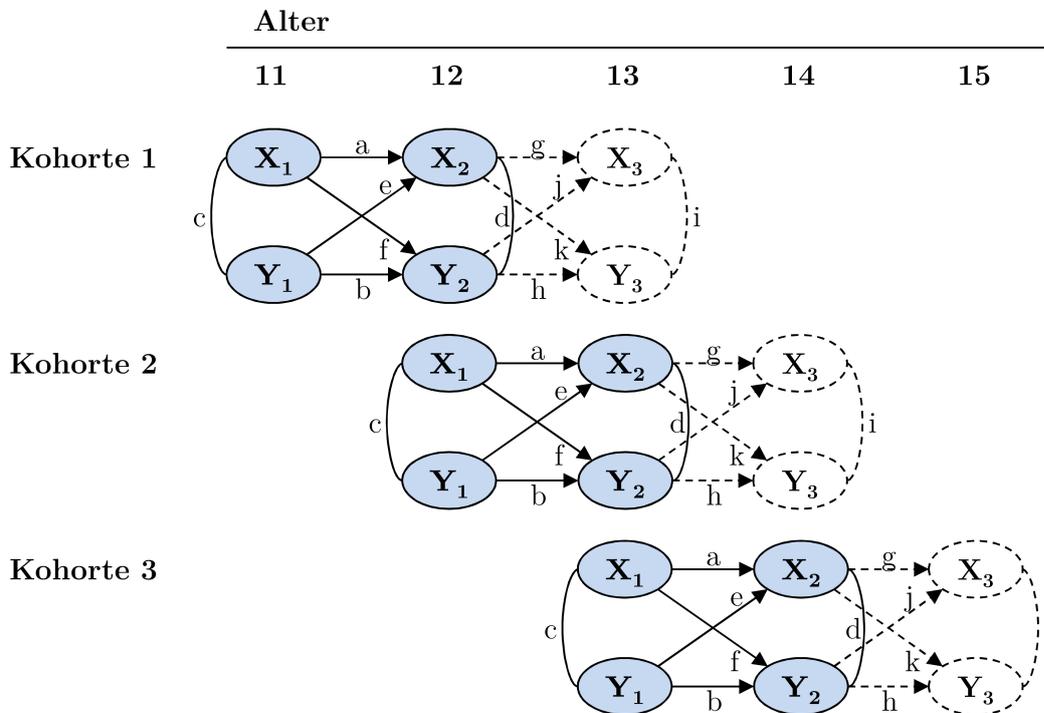


Abbildung 9.3. Prototypischen Multi-Kohorten-Sequenz-Design. Eigene Darstellung in Anlehnung an Abbildung 1 aus T. D. Little, Preacher, Selig und Card, 2007, S. 364.

Die Überprüfung von Kohorten- und Entwicklungseffekten erfolgt durch Testung verschiedener Submodelle gegen ein vollständiges Modell in welchem alle Pfade ($a - i$) zwischen den Kohorten und Zeitpunkten frei geschätzt werden. Zunächst wird ein Modell spezifiziert, in dem alle interessierenden Pfade (welche dies sind, hängt von den theoretischen Vorannahmen ab) zwischen den Kohorten und Zeitpunkten gleichgesetzt werden (=voll restringiertes Modell). Resultiert hieraus eine signifikante Verschlechterung des Modell-Fits im Vergleich zum vollständigen Modell, kann in der Folge abgeschätzt werden, in welchem Ausmaß diese Verringerung im Fit auf Kohorten- bzw. Entwicklungseffekte zurückgeht (vgl. zu diesem Vorgehen T. D. Little, Preacher, Selig & Card, 2007). Zur Bewertung des Kohorteneffektes wird die Erhöhung im χ^2 -Wert eines Modells mit gleichgesetzten interessierenden Pfaden zwischen den Kohorten im überlappenden Zeitintervall im Vergleich zum vollständigen Modell herangezogen. Ginge es beispielsweise um die Vorhersage von Y im Alter 13 durch X im Alter 12 wäre der CL-Pfad im überlappenden Zeitintervall wie folgt gleichzusetzen: X_2 nach Y_3 (k) in Kohorte 1 = X_1 nach Y_2 (f) in Kohorte 2. Die Verschlechterung des Modell-Fits, der

aus dieser Restriktion resultiert, bildet den Kohorteneffekt ab. Die verbleibende Veränderung im Modell-Fit des vollständigen im Vergleich zum voll restringierten Modell kann dann als Entwicklungseffekt interpretiert werden (T. D. Little et al., 2007).

In sämtlichen, zuvor beschriebenen Modellen gilt es, vor der vergleichenden Modellschätzung die *Invarianz der Messung* zu prüfen (z. B. Kline, 2011). Diese bezieht sich im Rahmen einer MKSD sowohl auf die Messäquivalenz über die Zeit als auch auf die Messäquivalenz zwischen den Kohorten. Für die Interpretierbarkeit der Ergebnisse ist von entscheidender Bedeutung, dass sich die psychometrischen Eigenschaften der wiederholt eingesetzten Inventare nicht verändert haben, bzw. sich nicht zwischen den Kohorten unterscheiden (Christ & Schlüter, 2012; Geiser, 2010). Die Bestätigung dieser Messinvarianz ist unmittelbar an die Interpretierbarkeit der Ergebnisse geknüpft. So wäre im Falle längsschnittlicher Analysen beispielsweise zu hinterfragen, ob in diesem Fall die Autokorrelation einer Variablen überhaupt als Stabilität interindividueller Unterschiede in demselben Konstrukt interpretiert werden dürfte (Geiser, 2010). In der Literatur werden unterschiedliche Stufen der Messinvarianz beschrieben (für eine ausführliche Darstellung vgl. z. B. Weiber & Mühlhaus, 2010), der Grad an Invarianz kann in Strukturgleichungsmodellen wiederum mit Hilfe von Gleichheitsrestriktionen und einem Vergleich der Modellpassung geprüft werden. Die schwächste Form der Invarianz stellt die *konfigurale Invarianz* dar. Diese liegt bereits dann vor, wenn sich die Faktorenstruktur (Anzahl der Faktoren und Ladungsstruktur) zwischen den Zeitpunkten und Gruppen nicht unterscheidet. Um diesem Kriterium zu genügen, muss keiner der Parameter exakt identisch sein (Geiser, 2010). Das Kriterium der *schwachen faktoriellen Invarianz* ist dann erfüllt, wenn zusätzlich zur konfiguralen Invarianz die Faktorladungen über die Zeit bzw. Gruppen hinweg konstant sind. Zur Prüfung dieser Bedingung wird ein Modell mit gleichgesetzten Faktorladungen zwischen Zeitpunkten und Gruppen spezifiziert und gegen das unrestringierte Modell getestet. *Starke faktorielle Invarianz* ist gegeben, wenn zusätzlich auch die Konstanten (Measurement Intercepts) der Indikatoren der Messmodelle übereinstimmen (Weiber & Mühlhaus, 2010). Die stärkste Form der Messinvarianz ist schließlich die *strikte faktorielle Invarianz* die vorliegt, wenn zusätzlich zur Gleichheit der Faktorladungen und Konstanten auch die Messfehlervarianzen invariant sind.

Generell gilt, dass mindestens das Kriterium der *konfiguralen Invarianz* erfüllt sein muss, um längsschnittliche Modelle bzw. vergleichende Betrachtungen zwischen Kohorten durchführen zu dürfen (Kline, 2011). Hingegen liegt *strikte faktorielle Invarianz* nur in den seltensten Fällen vor. Kann die Annahme der Messinvarianz lediglich für einen Teil der Indikatoren gehalten werden, liegt eine *partielle Messinvarianz* vor (Geiser, 2010).

10 Analysen der ätiologischen Betrachtung

Der Vergleich der Ähnlichkeit zwischen EZ und ZZ, wie in Kapitel 4 beschrieben, stellt einen ersten wichtigen Schritt beim Verständnis der Ätiologie verschiedener Merkmale dar. Allerdings trägt er Unterschieden in der Stichprobengröße von EZ und ZZ Paaren ebensowenig Rechnung wie Unterschieden in Mittelwert und Varianz der Gruppen (Riemann & F. M. Spinath, 2005). Umfassender kann die Schätzung der Parameter durch Anpassung von Strukturgleichungsmodellen vorgenommen werden (Rijsdijk & Sham, 2002). Dies bietet den Vorteil, dass durch die Schätzung der Effektstärke genetischer und Umwelteinflüsse die Vermutungen über die Höhe deren Einflusses auf interindividuelle Unterschiede geprüft werden können. Allgemein liegt auch in verhaltensgenetischen Analysen der Vorteil der SEM in der Möglichkeit, die absolute Passung eines Modells und die relative Passung von Submodellen zu testen (Y. L. L. Luo et al., 2011).

Nachfolgend wird zunächst das univariate Modell für Zwillingsdaten erläutert, anschließend wird die Erweiterung zum multivariaten Design beschrieben.

10.1 Univariate genetische Analysen

Die hier durchgeführten univariaten genetischen Analysen basieren auf den Kernannahmen des CTD (kein Vorliegen selektiver Partnerwahl, die Gültigkeit der EEA und die genetische Relation von EZ und ZZ), welche unter Abschnitt 4.4 bereits eingehend erläutert wurden. Ein univariates, latentes Pfaddiagramm für den Vergleich gemeinsam aufgewachsener EZ und ZZ ist in Abbildung 10.1 dargestellt (Neale & Cardon, 1992).

Das Modell (auch als *common-pathway model* bezeichnet) spezifiziert die Annahme, dass der Phänotyp von Zwilling 1 und Zwilling 2 auf die jeweiligen latenten additiven genetischen Effekte (A_X), Dominanzeffekte (D_X), Effekte der geteilten Umwelt (C_X) oder Effekte der nichtgeteilten Umwelt (E_X) zurückgeführt werden können. Da X und Y hier latente Variablen darstellen, werden zusätzlich für jede Variable (Var_{X1} und Var_{X2}) spezifische A , D , C und E Effekte angenommen, die die verbleibende Varianz erklären (Young, Stallings, Corley, Krauter & Hewitt, 2000). Im Gegensatz zur Betrachtung manifester Variablen können bei latenter Modellierung die nichtgeteilten

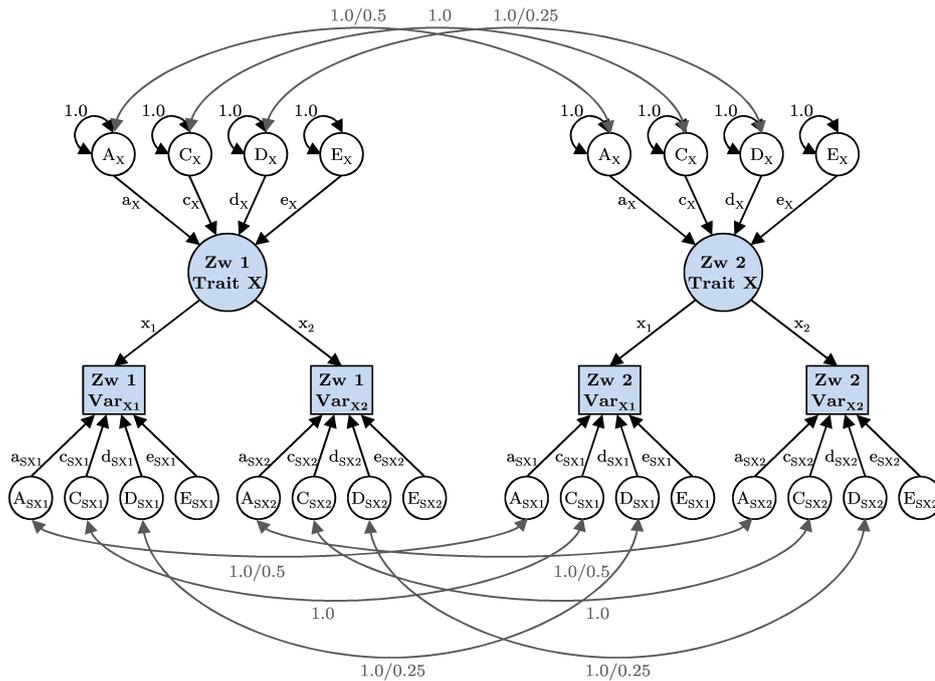


Abbildung 10.1. Univariates latentes Pfaddiagramm für den Vergleich gemeinsam aufgewachsener EZ und ZZ. *A* steht für die additiven genetischen Effekte, *C* für die geteilten Umwelteffekte, *D* für die Dominanzeffekte und *E* für die nichtgeteilten Umwelteffekte. Mit *a*, *c*, *d*, *e* sind die entsprechenden Pfadkoeffizienten bezeichnet. Die hellgrauen gebogenen Doppelpfeile geben die Ähnlichkeit innerhalb der EZ/ZZ Paare an.

Umwelteffekte als reine nichtgeteilte Umwelteffekte interpretiert werden, da der *spezifische E* Anteil in diesem Fall dem Messfehler¹ in dieser Art kontrolliert, fallen normalerweise die Schätzungen der Ererblichkeit etwas höher aus (Jang et al., 1998; Neale & Cardon, 1992). Außerdem wird festgesetzt, dass die Pfadkoeffizienten für Zwilling 1 und Zwilling 2 nicht als Funktion ihrer genetischen Relation variieren und keine reziproken Zwillingseffekte existieren (Neale & Cardon, 1992).

Innerhalb des CTD liegt, wie bereits angesprochen, eine Konfundierung von Dominanzeffekten und Effekten der geteilten Umwelt vor, da sich beide auf die relative Ähnlichkeit der ZZ auswirken (vgl. dazu Kapitel 4; Plomin et al., 2008). Folglich können beide Effekte in diesem Design nicht simultan geschätzt werden (Ozaki, Toyoda, Iwama, Kubo & Ando, 2011). Dies gelingt nur durch Hinzunahme weiterer Verwandtschaftsgruppen, die keine Dominanzeffekte teilen und damit die Trennung beider Effekte ermöglichen. Aus den ICCs kann abgeleitet werden, welcher der beiden Komponenten wahrscheinlicher ist und ob in der Folge als vollständiges Modell ein *ACE* oder *ADE* Modell geschätzt wird. Bei Ähnlichkeitskoeffizienten von ZZ, die weniger als halb so groß wie diejenigen der EZ sind, wird angenommen, dass die geteilte Umwelt keinen bedeutenden Einfluss auf die phänotypische Merkmalsausprägung hat und demzufolge Dominanz-

¹ Dieser kann auch aufgrund einer geringen Reliabilität der verwendeten Maße hoch sein.

effekte² angenommen werden können. Liegen jedoch ZZ Korrelationen vor, die höher als die Hälfte der EZ Korrelation ausfallen, impliziert dies, dass Effekte der geteilten Umwelt zur Merkmalsvarianz beitragen. Hier ist zu berücksichtigen, dass die Schätzung des ein oder anderen Modells nicht gleichbedeutend mit einem Nichtvorhandensein des anderen Einflusses ist (Neale, 2009).

Die Pfadkoeffizienten a , c (d) und e sowie der Mittelwert werden in diesem Modell als freie Parameter aus den empirischen Daten geschätzt. Alle anderen Parameter (die Varianz der latenten Variablen sowie deren Kovarianz) werden gemäß den Vorannahmen des CTD auf die entsprechenden Werte fixiert (Neale, 2009). Typischerweise wird angenommen, dass sich die Mittelwerte und Varianzen innerhalb der Paare und zwischen den EZ und ZZ Paaren nicht unterscheiden (Neale, 2009). Um die Passung des theoretischen Modells mit den beobachteten Daten zu beurteilen, wird auf die Varianz-Kovarianzmatrix zurückgegriffen. Abgeleitet aus Abbildung 10.1 ergeben sich für EZ bzw. ZZ folgende Matrizen:

$$\text{Cov } EZ = \begin{bmatrix} a^2 + d^2 + c^2 + e^2 & \\ a^2 + d^2 + c^2 & a^2 + d^2 + c^2 + e^2 \end{bmatrix}$$

$$\text{Cov } ZZ = \begin{bmatrix} a^2 + d^2 + c^2 + e^2 & \\ \frac{1}{2}a^2 + \frac{1}{4}d^2 + c^2 & a^2 + d^2 + c^2 + e^2 \end{bmatrix}$$

Die eigentliche Parameterschätzung erfolgt analog zu dem unter Kapitel 9 beschriebenen Vorgehen bei der phänotypischen Analyse von Strukturgleichungsmodellen. Auch hier wird eine möglichst gute Passung zwischen vorhergesagter und beobachteter Kovarianzmatrix angestrebt, wobei die Varianz-Kovarianzmatrix zwischen EZ und ZZ als Grundlage dient. Die Schätzung der freien Parameter erfolgt iterativ über die Minimierung der Differenz zwischen der beobachteten Kovarianzmatrix und der modellimplizierten Kovarianzmatrix (Reinecke, 2005). Im Rahmen vorliegender Arbeit wird auf das Computerprogramm Mx (Neale, Boker, Xie & Maes, 2003) zur Parameterschätzung zurückgegriffen.

In der univariaten Modellierung wird nun zunächst das vollständige *ACE* bzw. *ADE* Modell geschätzt und der Fit des Modells bewertet. Anschließend wird dieses gegen reduzierte Modelle (*CE*, *AE*, *E* bzw. *DE*, *AE*, *E*) getestet. Dementsprechend wird geprüft, welches dieser sparsameren Modelle nicht signifikant schlechter fittet als das Gesamtmodell und gleichzeitig einen noch akzeptablen Fit aufweist. Hierfür wird der LRT (siehe Abschnitt 9.2) herangezogen. Als weitere Kennwerte zur Absicherung der Güte

² EZ teilen alle nicht-additiven genetischen Effekte, ZZ im Schnitt lediglich 25% (vgl. 4.3.1).

der Modellschätzung werden in verhaltensgenetischen Analysen der AIC und RMSEA (vgl. Abschnitt 9.2) bewertet.

Darüber hinaus werden die Konfidenzintervalle (KI) der geschätzten Parameter betrachtet. Schließt das KI Null nicht ein, kann von der Signifikanz der Parameter ausgegangen werden (Neale & Cardon, 1992). Bei der Interpretation der errechneten Effekte muss allerdings immer bedacht werden, dass diese nicht zwangsläufig das wahre Verhältnis zwischen genetischen Effekten und Umwelteffekten beschreiben – hiervon kann nur dann ausgegangen werden, wenn das angenommene Modell auch tatsächlich passend ist. Haben beispielsweise Dominanzeffekte eigentlich einen Einfluss, werden aber nicht im Modell geschätzt, wird der Großteil an Varianz dieser nicht-additiven genetischen Effekte den additiven genetischen Effekten zugerechnet. Dies macht zwar eine Trennung beider genetischer Effekte unmöglich, führt aber nicht zu einem verzerrten Abbild der Ätiologie eines Merkmals. Wird andererseits ein *ADE* Modell spezifiziert, werden potenziell vorliegende Effekte der geteilten Umwelt ebenfalls dem additiven genetischen Effekt zugerechnet, was in einer Überschätzung des genetischen Einflusses und damit auch in einem verfälschten ätiologischen Muster resultiert (Y. L. L. Luo et al., 2011).

10.2 Multivariate genetische Analysen

Die bisherige Darstellung verhaltensgenetischer Analysen bezog sich lediglich auf genetische und umweltbedingte Einflüsse auf ein einzelnes phänotypisches Merkmal. Dieser Ansatz wird daher auch als univariater Ansatz bezeichnet. Das univariate Modell kann allerdings leicht zu einem multivariaten Modell erweitert werden, indem mehrere Merkmale (oder das gleiche Merkmal zu mehreren Messzeitpunkten) für jedes Individuum erfasst werden (Boomsma & Molenaar, 1987; Eaves & Gale, 1974; Posthuma, 2009). Während im Rahmen univariater genetischer Analysen die Varianz eines einzelnen Merkmals in genetische und umweltbedingte Anteile zerlegt wird, beziehen sich multivariate genetische Analysen auf die Kovarianz zwischen Merkmalen (Eaves & Gale, 1974). Dementsprechend liegt multivariaten genetischen Analysen die Frage nach den Ursachen (genetischen oder umweltbedingten) phänotypischer Korrelationen zwischen Merkmalen zugrunde (Kovas et al., 2007).

Hier spielt auch die Überlegung eine Rolle, dass multivariate genetische Analysen einen wichtigen Beitrag bei der Identifikation von Kandidatengenen für komplexe Verhaltensweisen darstellen: Bei der Betrachtung einzelner Merkmale finden sich relativ häufig genetische Einflüsse. Es ist jedoch recht unwahrscheinlich, dass jedes Merkmal von völlig unterschiedlichen Genen beeinflusst wird. Vielmehr können Gene einen Einfluss

auf verschiedenen Merkmale besitzen und auf diese Weise Korrelationen zwischen den Merkmalen bewirken (Plomin et al., 2008).

Als Informationsquelle dient im multivariaten Zwillingsdesign der Vergleich der Kreuzkorrelation (*Cross-Trait Cross-Twin Correlation*; CTCT) von EZ und ZZ. Eine CTCT ist eine Korrelation zwischen Merkmal X eines Zwillings und Merkmal Y seines Ko-Zwillings.³ Eine höhere Kreuzkorrelation der EZ im Vergleich zu ZZ impliziert, dass die beobachtete phänotypische Korrelation zwischen Merkmal X und Y von genetischen Faktoren abhängt (dies wird als *genetische Vermittlung* bzw. *bivariate Erbllichkeit* bezeichnet), während bei ähnlichen CTCT Korrelationen bei EZ und ZZ davon ausgegangen werden kann, dass die Kovariation der Merkmale auf (geteilte) Umweltfaktoren zurückzuführen ist (Posthuma, 2009; Riemann & F. M. Spinath, 2005). Entsprechend ergibt sich die folgende 2×2 Varianz-Kovarianz Matrix, die sich für mehr als zwei Merkmale zu einer $2n \times 2n$ Matrix, mit n gleich Anzahl der Variablen, erweitern lässt.

$$\begin{bmatrix} \text{Var}(X_1) & & & & \\ \text{Cov}(X_1, X_2) & \text{Var}(X_2) & & & \\ \text{Cov}(X_1, Y_1) & \text{Cov}(X_2, Y_1) & \text{Var}(Y_1) & & \\ \text{Cov}(X_1, Y_2) & \text{Cov}(X_2, Y_2) & \text{Cov}(Y_1, Y_2) & \text{Var}(Y_2) & \end{bmatrix}$$

In der Diagonale finden sich vier Varianzen, für jede Messung für Zwilling 1 und Zwilling 2. $\text{Cov}(X_1, Y_1)$ und $\text{Cov}(X_2, Y_2)$ sind die phänotypischen Kovarianzen zwischen X und Y für Zwilling 1 und Zwilling 2. $\text{Cov}(X_1, X_2)$ und $\text{Cov}(Y_1, Y_2)$ bezeichnen die univariaten Kovarianzen innerhalb der Paare und $\text{Cov}(X_1, Y_2)$ und $\text{Cov}(X_2, Y_1)$ sind schließlich die CTCT Kovarianzen. Eine Darstellung des aus dieser Varianz-Kovarianzmatrix resultierenden bivariaten Pfadmodells als *correlated factors model* findet sich in Anhang A.

Wie zuvor bereits erwähnt besteht auch die Möglichkeit der längsschnittlichen verhaltensgenetischen Analyse, die eine multivariate Analyse eines Merkmals zu verschiedenen Zeitpunkten darstellt. Die Analyse von Kontinuität und Veränderung kann in Rahmen verhaltensgenetischer Untersuchungen generell unter zweierlei Gesichtspunkten erfolgen: Zum einen können *quantitative Unterschiede* in der Höhe genetischer und Umwelteinflüsse betrachtet werden. Hierfür können Querschnittsdaten aus Stichproben in unterschiedlichen Altersbereichen herangezogen werden (Kovas et al., 2007). Zum anderen können *qualitative Veränderungen* in genetischen und Umwelteinflüssen über die Zeit durch Analyse genetisch sensitiver Längsschnittsdaten analysiert werden. Durch

³ Der bivariate Fall stellt dabei die einfachste Erweiterung der univariaten Betrachtung dar. Die Erweiterung auf mehr als zwei Merkmale kann in analoger Weise vorgenommen werden.

diese Zerlegung wird es möglich, Aussagen über genetische bzw. Umwelteinflüsse auf die Stabilität und Veränderung von Merkmalen zu treffen (Kovas et al., 2007; Neale, Boker, Bergeman & Maes, 2006; Plomin et al., 2008). Das in Abbildung 10.1 dargestellte Modell (die sogenannte *Cholesky Zerlegung*; Loehlin, 1996; Neale et al., 2003) testet die gemeinsamen und spezifischen genetischen und Umwelteffekte auf die Varianz in mehreren Merkmalen oder auf ein Merkmal zu unterschiedlichen Zeitpunkten.

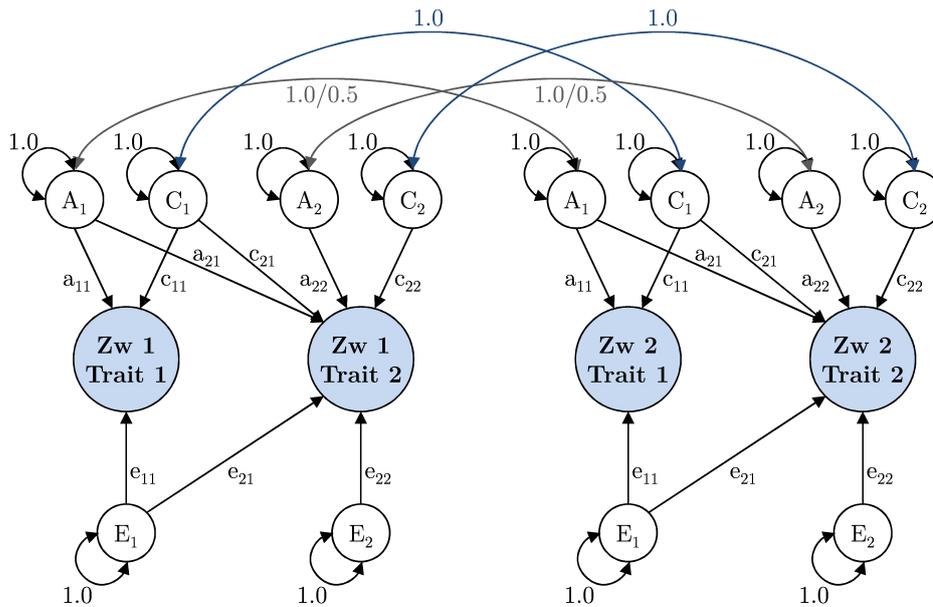


Abbildung 10.1. Bivariate latente Cholesky Zerlegung. *A* steht für die additiven genetischen Effekte, *C* für die geteilten Umwelteffekte und *E* für die nichtgeteilten Umwelteffekte. Mit *a*, *c*, *e* sind die entsprechenden Pfadkoeffizienten bezeichnet. Die gebogenen Doppelpfeile geben die Ähnlichkeit innerhalb der EZ/ZZ Paare an.

Die Cholesky Zerlegung entspricht einer phänotypischen hierarchischen Regression in welcher der inkrementelle Varianzanteil eines Prädiktors bestimmt wird (Kovas et al., 2007). In diesem Modell wird ebenfalls die Kovarianz zwischen Paaren von Variablen in eine genetische Komponente, sowie eine geteilte und nichtgeteilte Umweltkomponente zerlegt, wodurch eine Schätzung der Varianzanteile für jede Komponente sowie deren Zusammenhangs möglich wird (Plomin et al., 2008). Der jeweils erste latente genetische und geteilte sowie nichtgeteilte Umweltfaktor lädt auf allen beobachteten Variablen, der zweite auf allen Variablen außer der ersten usw. (Johnson et al., 2007). Wichtig ist hierbei, dass die Reihenfolge der beobachteten Variablen frei wählbar ist (Posthuma, 2009). Dies bedeutet, dass jede mögliche Anordnung der beobachteten Variablen die Varianz-Kovarianzmatrix in gleicher Weise erklärt (Loehlin, 1996). Es ist allerdings üblich, die beobachteten Variablen in ihrer angenommenen Wichtigkeit zur Erklärung der Outcome Variable zu ordnen und die Outcome Variable als letzte Variable einzuführen. Damit bilden die Pfadkoeffizienten, die von den latenten Varianzkomponenten zur be-

obachteten Outcome Variable führen, deren von den anderen Variablen unabhängigen Einfluss auf die Outcome Variable ab (Johnson et al., 2007). Eine Ausnahme von dieser Regel der Gleichheit in der Anordnung der beobachteten Variablen ergibt sich, wenn die beobachteten Merkmale einer natürlichen Ordnung folgen, also beispielsweise wenn Messungen der gleichen Variable zu verschiedenen Zeitpunkten vorliegen (Posthuma, 2009). Die gleiche Logik kann auch auf mehrere Variablen ausgeweitet werden.

Auf Grundlage dieses Modells können zwei weitere relevante Statistiken abgeleitet werden: Die *bivariate Erblichkeit* (bzw. die bivariate geteilte und nichtgeteilte Umwelt) sowie die *genetische Korrelation* (bzw. die genetische und Umweltkorrelation) zweier Merkmale.

Die bivariate Erblichkeit (biv_{a^2} ; bzw. die bivariate geteilte und nichtgeteilte Umwelt) gibt an, inwiefern phänotypisch beobachtete Korrelationen zwischen Merkmalen durch genetische (bzw. geteilte und nichtgeteilte Umwelt-) Prozesse vermittelt sind (Plomin & F. M. Spinath, 2002; Plomin et al., 2008). Die Höhe der bivariaten genetischen und Umwelteinflüsse sind dabei unabhängig von der Höhe der beobachteten phänotypischen Korrelation und der Höhe der genetischen Effekte und Umwelteffekte auf ein Merkmal (Johnson et al., 2007). Sie geben den Anteil an, zu dem die phänotypische Korrelation auf genetische und Umwelteffekte zurückgeht und addieren sich folglich zu 1.00 (Johnson et al., 2007). Für die bivariate Erblichkeit zweier Merkmale ergibt sich an Hand der Pfade von Abbildung 10.1 nachfolgende Gleichung, die auf die Berechnung der bivariaten Umwelteinflüsse übertragen werden kann:

$$biv_{a^2} = \frac{a_{11}a_{21}}{a_{11}a_{21} + c_{11}c_{21} + e_{11}e_{21}}$$

Liegen mehr als zwei Variablen vor, muss bei der Berechnung der bivariaten Erblichkeit zwischen Merkmal 2 und Merkmal 3 der Einfluss von Merkmal 1 auf Merkmal 2 bei der Berechnung entsprechend berücksichtigt werden.

Die *genetische Korrelation* (r_A) bezeichnet das Ausmaß, in dem genetisch bedingte Prozesse, die auf ein Merkmal wirken, mit den genetischen Prozessen, die auf ein anderes Merkmal wirken, überlappen (Plomin et al., 2008). Gleiches gilt für die Korrelationen zwischen den geteilten und nichtgeteilten Umwelteinflüssen. Die genetische Korrelation (bzw. die geteilte (r_C) und nichtgeteilte (r_E) Umweltkorrelation) kann Werte zwischen -1.00 und 1.00 annehmen. Ein Wert von 1.00 würde implizieren, dass Merkmal X und Y sämtliche additiven genetischen Einflüsse teilen. Eine geteilte Umweltkorrelation von 0.00 legt hingegen nahe, dass die Umwelteinflüsse die Zwillinge im Merkmal X einander ähnlicher machen, unabhängig sind von den Umwelteinflüssen, die Zwillinge einander im Merkmal Y ähnlicher machen (Plomin et al., 2008). Die genetische

Korrelation zwischen Merkmal X und Merkmal Y berechnet sich als

$$r_A = \frac{a_{21}}{\sqrt{a_{21}^2 + a_{22}^2}}$$

Analog gilt für die geteilte und nichtgeteilte Umweltkorrelation:

$$r_C = \frac{c_{21}}{\sqrt{c_{21}^2 + c_{22}^2}} \quad \text{bzw.} \quad r_E = \frac{e_{21}}{\sqrt{e_{21}^2 + e_{22}^2}}$$

Die genetische Korrelation sowie die geteilte Umwelt und nichtgeteilte Umweltkorrelation sind dabei unabhängig sowohl von der Höhe der univariat ermittelten genetischen und Umwelteffekte als auch von der Höhe der phänotypischen Korrelation. Auch wenn zwei Merkmale eine geringe Erblichkeit aufweisen, kann die genetische Korrelation zwischen ihnen hoch sein. Obwohl beide Merkmale dann für sich betrachtet lediglich geringer genetischer Beeinflussung unterliegen, würde eine hohe genetische Korrelation zwischen Merkmal X und Y bedeuten, dass, welches Gen auch immer Merkmal X beeinflusst, dieses Gen wahrscheinlich auch einen Einfluss auf Merkmal Y hat. Der genetischen Korrelation wird eine hohe praktische Relevanz bei der Suche nach Kandidatengenen zugesprochen (vgl. z. B. Posthuma, 2009). Liegt eine hohe genetische Korrelation vor ($r_A > 0.50$), ist die Wahrscheinlichkeit, dass ein identifiziertes Gen für ein Merkmal auch das andere Merkmal beeinflusst, als relativ hoch einzuschätzen (vgl. z. B. Plomin et al., 2008). Diese Interpretation genetischer Korrelation als *Pleiotropie* stellt jedoch nur eine Möglichkeit zur Erklärung einer genetischen Korrelation dar. Analog zur Interpretation phänotypischer Korrelationen würde sich auch dann eine genetische Korrelation zweier Merkmale finden, wenn ein genetisch beeinflusstes Merkmal X ein Merkmal Y beeinflusst oder umgekehrt (Johnson et al., 2011). Genauso könnten die drei Mechanismen auch gleichzeitig wirken (De Geus, 2011). Die Analyse von r_A , r_C und r_E geht folglich über die Beantwortung der Frage, *ob* bestimmte Merkmale miteinander zusammenhängen hinaus, und liefert Hinweise auf die Frage, *warum* diese Merkmale miteinander korrelieren (Plomin et al., 2008).

Die Anwendung der Cholesky Zerlegung auf CLP-Analysen macht es nun möglich, das Ausmaß, zu dem die phänotypisch gefundenen CL-Beziehungen genetisch bzw. umweltbedingt vermittelt sind, zu bestimmen (Y. L. L. Luo, Haworth & Plomin, 2010).

Das in Abbildung 10.2 gezeigte *genetische CL-Modell* enthält drei Komponenten (Y. L. L. Luo et al., 2010): Eine multivariate Analyse zwischen X und Y zu jedem Zeitpunkt, eine längsschnittliche Analyse von X und Y über das betrachtete Zeitinter-

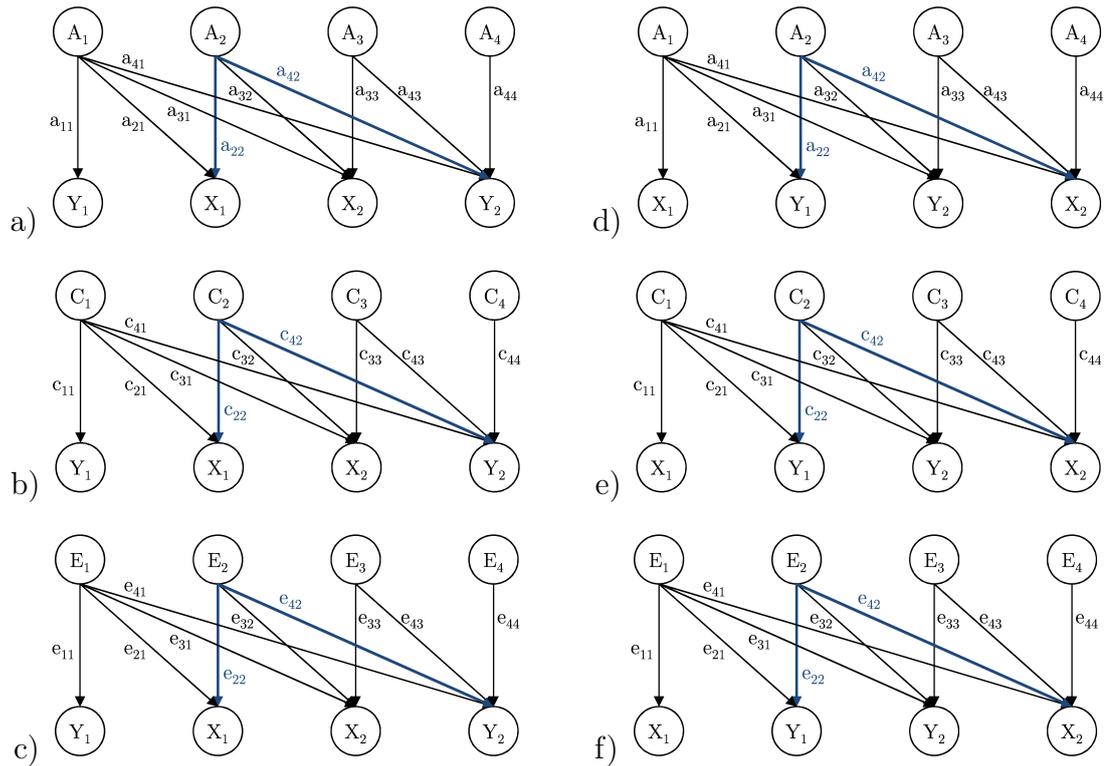


Abbildung 10.2. Genetisches Cross-Lagged Modell. Eigene Darstellung in Anlehnung an Abbildung 1 Y. L. L. Luo et al., 2010, S. 431. Teil a–c stellt die Zerlegung des CL-Pfades von X_1 nach Y_2 dar, e–f die Zerlegung des CL-Pfades von Y_1 nach X_2 . A steht für die additiven genetischen Effekte, C für die geteilten Umwelteffekte und E für die nichtgeteilten Umwelteffekte. Mit a, c, e sind die entsprechenden Pfadkoeffizienten bezeichnet. Auf die Darstellung der spezifischen A, C , und E Komponenten wurde hier zugunsten einer übersichtlicheren Darstellung verzichtet.

vall, und schließlich die Analyse der CL-Beziehung von X und Y .

Wie Abbildung 10.2 zu entnehmen ist, wird die Schätzung dieser Komponenten durch eine spezifische Anordnung der Variablen möglich: Abbildung 10.2 a–c stellt die Zerlegung des CL-Pfades von X_1 nach Y_2 dar, e–f die Zerlegung des CL-Pfades von Y_1 nach X_2 . Das Modell untersucht dementsprechend immer die Beziehung der Variablen an der zweiten und vierten Position und kontrolliert (analog zum phänotypischen CL-Modell) um konfundierende Einflüsse der Stabilität sowie der zeitgleichen Korrelation von X und Y . Das Modell beinhaltet insgesamt 12 latente Faktoren, $A_1 - A_4$, $C_1 - C_4$ und $E_1 - E_4$. Die jeweiligen Pfadschätzungen können zur Berechnung der (phänotypischen) Beziehungen herangezogen werden (hier am Beispiel a–c illustriert):

Zeitgleiche Beziehung Zeitpunkt 1 (Y_1 mit X_1):	$a_{11}a_{21} + c_{11}c_{21} + e_{11}e_{21}$
Umgekehrter CL-Pfad (Korrelation Y_1 mit X_2):	$a_{11}a_{31} + c_{11}c_{31} + e_{11}e_{31}$
Stabilität von Y :	$a_{11}a_{41} + c_{11}c_{41} + e_{11}e_{41}$
Stabilität von X :	$a_{22}a_{32} + c_{22}c_{32} + e_{22}e_{32}$
Interessierender CL-Pfad X_1 nach Y_2 :	$a_{22}a_{42} + c_{22}c_{42} + e_{22}e_{42}$
Zeitgleiche Beziehung Zeitpunkt 2 (Y_2 mit X_2):	$a_{33}a_{43} + c_{33}c_{43} + e_{33}e_{43}$
Spezifische A, C, E Schätzung von Y_2 :	$a_{44} + c_{44} + e_{44}$

Der genetische bzw. Umweltanteil an der phänotypischen CL-Korrelation kann dann als relativer Anteil bestimmt werden, also beispielsweise für den genetischen Anteil als (Y. L. L. Luo et al., 2010):

$$\frac{a_{22}a_{42}}{a_{22}a_{42} + c_{22}c_{42} + e_{22}e_{42}}$$

Ein Einschränkung dieses Designs ist zu bedenken, dass nicht beide CL-Pfade simultan geschätzt werden können (Y. L. L. Luo et al., 2010).

Abschließend sei noch darauf hingewiesen, dass für Zwillingsanalysen in der Regel alters- und geschlechtskorrigierte standardisierte Residuen verwendet werden (Davis et al., 2007). Dies begründet sich aus dem Umstand heraus, dass das Alter der Zwillinge innerhalb eines Paares zu 1.00 korreliert. Bei Vorliegen von Alterseffekten würde dies dazu führen, dass auch der Alterseffekt zur Ähnlichkeit der Paare beitragen würde, was innerhalb der SEM Schätzung dann als Einfluss der geteilten Umwelt fehlinterpretiert werden würde (Eaves, Eysenck & Martin, 1989).

V. ERGEBNISSE

Die Darstellung der Ergebnisse erfolgt in vier Unterkapiteln. Zunächst werden die Lage-, Streuungs- und Verteilungsmaße sowie die Reliabilität der eingesetzten Instrumente für jede Kohorten und die Gesamtstichprobe berichtet (Kapitel 11). Ferner erfolgt in diesem ersten Teil die Betrachtung der manifesten Skaleninterkorrelationen. Im Anschluss an die deskriptive Betrachtung erfolgt die Prüfung der unter Kapitel 9.1 dargestellten Voraussetzungen zur Durchführung einer SEM und die Güteprüfung der latenten Modellierung der betrachteten Skalen (Kapitel 12). Hiernach werden die Ergebnisse der durchgeführten Analysen zur Beantwortung der phänotypischen (Kapitel 13) und ätiologischen (Kapitel 14) Fragestellungen berichtet.

11 Deskriptive Statistik und manifeste Skaleninterkorrelationen

11.1 Schulerfolg: Kennwerte und deskriptive Statistik

Im Folgenden wird zunächst, ergänzend zur allgemeinen Darstellung in Kapitel 8, konkretisiert, auf welche Schulnoten zur Operationalisierung des Schulerfolges zurückgegriffen wird. Anschließend erfolgt die Betrachtung der deskriptiven Kennwerte der verwendeten Schulnoten.

In KoSMoS liegen insgesamt die Schulnoten aus sechs Schuljahren vor, wobei für das letzte erfasste Schuljahr 2009/10 lediglich die Halbjahresnoten erfragt werden konnten. Analog zur Darstellung in Abbildung 7.5 und des unter Kapitel 9.3 beschriebenen Vorgehens der Analyse im MKSD werden die Noten aus den Schuljahren 2004/05, 2006/07 sowie 2008/09 zur Beantwortung der Fragestellungen herangezogen. Zu berücksichtigen ist an dieser Stelle, dass in den meisten Schulen erst zum Ende der zweiten Grundschulklasse erste Schulnoten vergeben werden, wodurch für K1 im Schuljahr 2004/05 in der Mehrzahl der Fälle lediglich die Jahresendnoten zur Verfügung standen. Die Noten wurden aus Gründen einer einfacheren Interpretierbarkeit rekodiert, so dass höhere Werte einer besseren Leistung entsprachen.

Tabelle 11.1 liefert einen Überblick über die relevanten deskriptiven Maße der jeweiligen Halbjahres- und Jahresnoten. Die Mittelwerte der Fachnoten lagen sich zwischen 4.11 (SD = 0.95; K3, JZ 2008/09) und 5.02 (SD = 0.79; K1, HZ 2004/05) im Fach Mathematik und zwischen 4.20 (SD = 0.75; K3, HZ 2008/09) und 4.76 (SD = 0.82; K1, HZ 2006/07) im Fach Deutsch. Die Mittelwerte lagen dabei in der Regel in den jüngeren Kohorten höher als in den älteren. Eine Ausnahme hiervon lag im Schulfach Mathematik im Schuljahr 2008/2009 vor, hier fielen die Mittelwerte von K2 jeweils geringfügig höher aus als in K1 (K2: M = 4.29, SD = 0.99 und M = 4.38, SD = 0.97 vs. K1: M = 4.24, SD = 1.00 und M = 4.36, SD = 0.91). In Deutsch überstiegen im Schuljahr 2004/2005 sowohl die Werte von K2 als auch diejenigen von K3 die Mittelwerte von K1 leicht (K1: M = 4.49, SD = 1.00 und M = 4.65, SD = 1.00 vs. K2: M = 4.73, SD = 0.80 und M = 4.68, SD = 0.77 vs. K3: M = 4.64, SD = 0.83 und M = 4.73,

SD = 0.76). In beiden Schulfächern sanken die Mittelwerte in allen Kohorten über die Zeit.

Die einzige Ausnahme hiervon fand sich in K1 in den Schulleistungen im Fach Deutsch. Hier stieg der Mittelwert von 2004/05 auf 2005/07 leicht an. Im Fach Mathematik konnte insgesamt ein stärkeres Absinken der Mittelwerte beobachtet werden.

Im Hinblick auf die Normalverteilungsannahme kann für beide Schulfächer von einer Nichtverletzung in allen Kohorten und Schuljahren ausgegangen werden: Alle Werte von Schiefe und Exzess lagen deutlich unterhalb der von West et al. (1995) angegebenen Grenzen. In der Tendenz lag in allen Kohorten sowohl in Mathematik als auch in Deutsch eine leicht linksschiefe, etwas breitgipfligere Verteilung vor.

Aus Tabelle 11.2 sind schließlich die Interkorrelationen der Schulnoten in Mathematik und Deutsch innerhalb und zwischen den Schuljahren sowie deren Reliabilitäten in den einzelnen Kohorten abgetragen. Die Angaben beziehen sich jeweils auf eine aus Halbjahres- und Jahresendnote gemittelte Gesamtnote.

Die fachinternen Korrelationen lagen im Fach Mathematik zwischen $r = 0.32$ und $r = 0.66$ sowie zwischen $r = 0.42$ und $r = 0.70$ im Fach Deutsch und damit in einem mittleren bis hohen Bereich (Cohen, 1988). Die Korrelationen fielen zwischen aufeinanderfolgenden Schuljahren in der Regel höher aus. Die diskriminanten Korrelationen innerhalb eines jeden Schuljahres lagen ebenfalls in einem mittleren Bereich ($r = 0.43$ bis $r = 0.62$) und auch über die verschiedenen Schuljahre hinweg lagen substantielle Korrelationen zwischen den Mathematik- und Deutschnoten vor (zwischen $r = 0.10$ und $r = 0.54$). Die internen Konsistenzen für Mathematik und Deutsch in sämtlichen Schuljahren und in allen Kohorten lag in einem hohen bis sehr hohen Bereich, tendenziell lag die Reliabilität dabei für Deutsch noch etwas höher als in Mathematik und in den Grundschuljahren höher als in der weiterführenden Schule ($\alpha_M = 0.80 - 0.95$; $\alpha_D = 0.80 - 0.97$).

Tabelle 11.1
Deskriptive Kennwerte und Verteilungsmaße der Schulnoten

	Mittelwert												Standardabweichung											
	2004/05				2006/07				2008/09				2004/05				2006/07				2008/09			
	HZ	JZ	HZ	JZ	HZ	JZ	HZ	JZ	HZ	JZ	HZ	JZ	HZ	JZ	HZ	JZ	HZ	JZ	HZ	JZ	HZ	JZ		
Ma	K_AS_2	4.80	4.77	4.56	4.58	4.23	4.29	0.80	0.81	0.84	0.88	0.80	0.81	0.84	0.87	0.88	0.80	0.81	0.84	0.87	0.89	0.95		
	K1_AS_2	5.02	4.95	4.76	4.76	4.24	4.36	0.79	0.90	0.76	0.81	0.79	0.90	0.76	0.81	0.81	0.79	0.90	0.76	0.81	1.00	0.91		
	K2_AS_2	4.82	4.82	4.62	4.64	4.29	4.38	0.80	0.82	0.81	0.87	0.80	0.82	0.81	0.87	0.87	0.80	0.82	0.81	0.87	0.99	0.97		
	K3_AS_2	4.71	4.67	4.32	4.34	4.14	4.11	0.80	0.78	0.87	0.90	0.80	0.78	0.87	0.90	0.90	0.80	0.78	0.87	0.96	0.96	0.95		
De	K_AS_2	4.66	4.70	4.50	4.46	4.31	4.30	0.84	0.81	0.79	0.82	0.84	0.81	0.79	0.82	0.82	0.84	0.81	0.79	0.81	0.81	0.81		
	K1_AS_2	4.49	4.65	4.76	4.74	4.38	4.43	1.00	1.00	0.82	0.81	1.00	1.00	0.82	0.81	0.81	1.00	0.82	0.81	0.81	0.81	0.86		
	K2_AS_2	4.73	4.68	4.44	4.41	4.35	4.32	0.80	0.77	0.70	0.83	0.80	0.77	0.70	0.83	0.83	0.80	0.77	0.70	0.83	0.83	0.81		
	K3_AS_2	4.64	4.73	4.34	4.26	4.20	4.24	0.83	0.76	0.79	0.77	0.83	0.76	0.79	0.77	0.77	0.83	0.76	0.79	0.75	0.75	0.76		
	Schiefe												Exzess											
	2004/05				2006/07				2008/09				2004/05				2006/07				2008/09			
	HZ	JZ	HZ	JZ	HZ	JZ	HZ	JZ	HZ	JZ	HZ	JZ	HZ	JZ	HZ	JZ	HZ	JZ	HZ	JZ	HZ	JZ	HZ	JZ
Ma	K_AS_2	-0.49	-0.44	-0.21	-0.24	-0.10	-0.12	-0.45	-0.34	-0.08	-0.10	-0.45	-0.34	-0.08	-0.10	-0.07	-0.45	-0.34	-0.08	-0.10	-0.10	-0.07		
	K1_AS_2	-0.92	-0.84	-0.23	-0.52	0.12	-0.06	1.17	0.30	-0.21	-0.02	-0.06	0.30	-0.21	-0.02	-0.20	1.17	0.30	-0.21	-0.02	-0.66	-0.20		
	K2_AS_2	-0.57	-0.50	-0.17	-0.23	-0.15	-0.17	0.40	0.20	-0.43	-0.31	-0.17	0.20	-0.43	-0.31	-0.46	0.40	0.20	-0.43	-0.40	-0.40	-0.46		
	K3_AS_2	-0.35	-0.35	-0.09	0.12	-0.29	-0.11	-0.20	-0.17	-0.26	-0.30	-0.11	-0.20	-0.17	-0.26	-0.63	-0.20	-0.17	-0.26	-0.12	-0.12	-0.63		
De	K_AS_2	0.10	-0.03	-0.29	-0.35	-0.41	-0.45	-0.04	-0.28	-0.27	-0.32	-0.04	-0.28	-0.27	-0.32	-0.24	-0.04	-0.28	-0.27	-0.32	-0.32	-0.24		
	K1_AS_2	-0.44	-0.35	-0.32	-0.32	-0.06	-0.09	-0.55	-1.03	-0.32	-0.26	-0.09	-0.55	-1.03	-0.26	-0.39	-0.55	-1.03	-0.32	-0.02	-0.02	-0.39		
	K2_AS_2	-0.48	-0.53	0.23	-0.01	0.30	0.22	-0.04	0.07	-0.13	-0.55	0.30	-0.04	0.07	-0.13	-0.36	-0.04	0.07	-0.13	-0.39	-0.39	-0.36		
	K3_AS_2	-0.32	-0.08	-0.26	-0.13	-0.10	-0.04	0.02	-0.43	-0.28	0.27	-0.04	-0.43	-0.28	0.27	0.17	0.02	-0.43	-0.28	-0.71	-0.71	0.17		

Anmerkungen. N(K_AS_2) = 318-442; N(K1_AS_2) = 38-128; N(K2_AS_2) = 135-200; N(K3_AS_2) = 117-183; HZ = Halbjahreszeugnis; JZ = Jahreszeugnis; Ma = Mathematik; De = Deutsch

Tabelle 11.2

Interkorrelation und Reliabilität der Schulnoten (Mathematik und Deutsch) innerhalb und zwischen den Schuljahren

		Mathematik			Deutsch			
		Schuljahr	04/05	06/07	08/09	04/05	06/07	08/09
K_AS_2	Ma	04/05	0.88	0.46	0.44	0.56	0.43	0.45
		06/07		0.83	0.60	0.23	0.52	0.45
		08/09			0.85	0.14*	0.37	0.51
	De	04/05				0.92	0.48	0.50
		06/07					0.86	0.62
		08/09						0.83
K1_AS_2	Ma	04/05	0.95	0.38*	0.33*	0.43	0.38*	0.48
		06/07		0.80	0.45	0.51	0.59	0.36
		08/09			0.86	0.10 ^{n.s.}	0.23*	0.49
	De	04/05				0.97	0.75	0.57
		06/07					0.89	0.53
		08/09						0.89
K2_AS_2	Ma	04/05	0.89	0.49	0.57	0.62	0.47	0.54
		06/07		0.84	0.65	0.16*	0.50	0.51
		08/09			0.85	0.14 ^{n.s.}	0.46	0.54
	De	04/05				0.93	0.45	0.51
		06/07					0.83	0.70
		08/09						0.80
K3_AS_2	Ma	04/05	0.85	0.41	0.32	0.60	0.38	0.35
		06/07		0.82	0.66	0.24*	0.43	0.45
		08/09			0.83	0.13 ^{n.s.}	0.37	0.48
	De	04/05				0.89	0.42	0.45
		06/07					0.83	0.63
		08/09						0.80

Anmerkungen. N(K_AS_2) = 215-442; N(K1_AS_2) = 35-128; N(K2_AS_2) = 92-169; N(K3_AS_2) = 88-145; Alle Korrelationen signifikant mit $p < 0.01$, es sei denn explizit anders gekennzeichnet; Reliabilität (Cronbachs α) = Werte in der Hauptdiagonale (fett); Diskriminante Korrelation zwischen den Fächern innerhalb eines Schuljahres = Nebendiagonale (fett).

* $p < 0.05$. ^{n.s.} nicht signifikant.

11.2 Allgemeine kognitive Fähigkeiten: Kennwerte und deskriptive Statistik

Wie unter 8.3 bereits beschrieben wurden die Kinder insgesamt zweimal hinsichtlich ihrer kognitiven Fähigkeiten getestet, wobei zum dritten Messzeitpunkt eine altersspezifische Erfassung im Rahmen der Telefontestung erfolgte. Den Kindern wurden hierzu für drei der vier Subtests (Figurenklassifikation, Figurenalogien, Wortanalogien) die der jeweiligen Klassenstufe entsprechenden Items vorgegeben. Für den zweiten ver-

balen Subtest (Allgemeinwissen) wurden hingegen für alle Kinder die gleichen Items eingesetzt. An der Telefon-Testung nahmen insgesamt 22 Kinder nicht mehr teil.

Aufgrund der unterschiedlichen Anzahl an Items im Untertest *Allgemeinwissen* zu Z1 (18 Items) und Z2 (21 Items; siehe auch 8.3.2), wurden die Skalen für die nachfolgende Betrachtung der deskriptiven Kennwerte an der Itemzahl der Subtests relativiert, um eine Vergleichbarkeit zwischen den einzelnen Tests, sowie zwischen den beiden Messungen zu ermöglichen. Die Skalenwerte liegen dementsprechend zwischen 0 und 1, da für jedes richtig gelöste Item ein Punkt gezählt wurde.

Die Verteilung der Mittelwerte lieferte für beide Zeitpunkte ein korrespondierendes Muster (vgl. Tabelle 11.3). Zum einen lag die mittlere Ausprägung in drei der vier Untertests (Figurenklassifikation, Wortanalogien und Allgemeinwissen) in allen Kohorten zu beiden Zeitpunkten in einem hohen Bereich. Insbesondere in der Skala Figurenklassifikation (NV1) fand sich in der zweiten Messung eine insgesamt hohe Ausprägung ($M = 0.76$; $SD = 0.16$ für K3 bis $M = 0.82$; $SD = 0.15$ für K1). Im Subtest Figurenanalogien (NV2) fanden sich die niedrigsten Mittelwerte, diese lagen zu beiden Messungen in einem mittleren Bereich ($M = 0.43$; $SD = 0.13$ für K1, Z1 bis $M = 0.59$; $SD = 0.13$ für K3, Z3). Zum anderen zeigte sich in der ersten Messung für die jüngsten Kinder die jeweils niedrigste Ausprägung pro Subtest. In der zweiten, altersspezifischen Erfassung erreichten die Kinder von K1 im Subtest NV1 die höchsten, in den Subtests NV2 sowie Allgemeinwissen (V2) die niedrigsten Werte. Im Untertest Wortanalogien (V1) lagen alle Kohorten gleich auf. Insgesamt betrachtet fielen die Mittelwerte in beiden nonverbalen Untertests in allen Kohorten zu Z3 etwas höher aus als zu Z1. In den beiden verbalen Skalen zeigte sich hingegen ein für die Kohorten unterschiedlicher Verlauf der Mittelwerte. Die Werte von K1 stiegen in beiden Skalen an, während die Werte der beiden älteren Kohorten von Z1 nach Z3 sanken.

Tabelle 11.3 informiert neben den Mittelwerten über die Verteilungsmaße der kognitiven Skalen. Grundsätzlich ließ sich in der zweiten Messung der kognitiven Fähigkeiten eine leichte Abweichung von der Normalverteilung feststellen.

Sämtliche Skalen wiesen zu beiden Zeitpunkten über alle Kohorten hinweg eine leichte Linksschiefe auf, Werte im oberen Bereich der Verteilung waren somit überrepräsentiert. Am stärksten fiel dies in allen Kohorten im Untertest NV1 aus ($SCH = -0.54$ bis $SCH = -0.95$ zu Z1; $SCH = -1.12$ bis $SCH = -1.92$ zu Z3). Eine nahezu symmetrische Verteilung fand sich für K1 zu Z1 und für K3 zu Z3 im Subtest V1 ($SCH = -0.01$ respektive $SCH = 0.02$). Bis auf wenige Ausnahmen lag zu Z1 in allen Untertests und für alle Kohorten mit Werten zwischen $EX = -0.73$ und $EX = -0.14$ eine etwas breitgipfligere Verteilung vor. Dementgegen zeigte sich in der kognitiven Messung zu Z3 eine schmalgipflige Verteilung der Skala NV1 und eine breitgipflige Verteilung der Skala V1

Tabelle 11.3

Deskriptive Kennwerte und Verteilungsmaße der kognitiven Skalen

		Mittelwert		Standardabweichung		Schiefe		Exzess	
		Z 1	Z 3	Z 1	Z 3	Z 1	Z 3	Z 1	Z 3
NV1	K_AS_2	0.72	0.80	0.18	0.15	-0.82	-1.44	-0.03	2.82
	K1_AS_2	0.65	0.82	0.19	0.15	-0.54	-1.92	-0.38	5.40
	K2_AS_2	0.74	0.81	0.17	0.13	-0.95	-1.31	0.39	2.11
	K3_AS_2	0.77	0.76	0.17	0.16	-0.93	-1.12	0.13	1.49
NV2	K_AS_2	0.49	0.57	0.13	0.13	-0.40	-0.79	-0.36	0.63
	K1_AS_2	0.43	0.52	0.13	0.14	-0.14	-0.42	-0.63	-0.34
	K2_AS_2	0.50	0.58	0.11	0.12	-0.14	-0.95	-0.73	1.76
	K3_AS_2	0.56	0.59	0.11	0.13	-0.51	-0.94	-0.28	1.36
V1	K_AS_2	0.65	0.62	0.18	0.15	-0.54	-0.18	-0.23	-0.43
	K1_AS_2	0.55	0.62	0.19	0.17	-0.01	-0.28	-0.14	-0.42
	K2_AS_2	0.67	0.63	0.16	0.14	-0.45	-0.16	-0.32	-0.51
	K3_AS_2	0.72	0.62	0.16	0.14	-1.02	0.02	0.99	-0.61
V2	K_AS_2	0.69	0.68	0.17	0.16	-0.40	-0.79	-0.36	0.63
	K1_AS_2	0.59	0.62	0.18	0.17	-0.14	-0.42	-0.63	-0.34
	K2_AS_2	0.70	0.69	0.15	0.14	-0.14	-0.95	-0.73	1.76
	K3_AS_2	0.76	0.70	0.15	0.15	-0.51	-0.94	-0.28	1.34

Anmerkungen. N(K_AS_2) = 370-575; N(K1_AS_2) = 111-170; N(K2_AS_2) = 139-215; N(K3_AS_2) = 120-190; NV1 = Figurenklassifikation; NV2 = Figurenanalogien; V1 = Wortanalogien; V2 = Allgemeinwissen; Die Skalen sind an der jeweiligen Itemzahl zu den Zeitpunkten relativiert (sämtliche Skalen je Zeitpunkt 25 Items, V2 zu Z1 18 Items, zu Z3 21 Items).

in allen Kohorten (EX = 1.49 bis EX = 5.40 bzw. EX = -0.61 bis EX = -0.42). In den beiden verbleibenden Skalen lag in K1 jeweils eine breitgipflige und in K2 sowie K3 eine schmalgipflige Verteilung vor. Sämtliche Werte von Schiefe und Exzess lagen demnach noch in den von West et al. (1995) angegebenen Grenzen. Aufgrund dessen und des Umstandes, dass lediglich einzelne Werte eine stärkere Abweichung anzeigten (NV1 zu Z3, K1) wurde in der Folge insgesamt von einer Nichtverletzung der Normalverteilungsannahme ausgegangen.

Die Inspektion der Verteilungsmaße legte die Annahme potenzieller Deckeneffekte vor allem in den nonverbalen Subtests nahe. Um diesen Aspekt einer genaueren Betrachtung zu unterziehen, wurde zusätzlich der Range der Skalen betrachtet. Hier zeigte sich für die Subskala NV1 zu Z1, dass 8 Kinder alle Aufgaben richtig gelöst haben, während es kein Kind gab, welches keine der Aufgaben lösen konnte. Insgesamt erhielten über 87% aller Kinder mindestens die Hälfte der erreichbaren Punktzahl. Noch extremer fiel dies in der zweiten Messung aus. 15 Kinder lösten alle Aufgaben richtig, zudem gab es kein Kind, das nicht mindestens drei Aufgaben richtig löste. Über 95% erreichten min-

destens die Hälfte der Punktzahl. Auch in der zweiten nonverbalen Skala zeichnete sich dieser Trend ab. Insbesondere zu Z1 lag der Anteil an Kindern mit voller Punktzahl sehr hoch ($n = 77$ entspricht 13,5%), ca. 87% der Kinder lösten über die Hälfte der Aufgaben. In der zweiten Messung lag dieser Anteil bei 83,5%, drei Kinder lösten alle Aufgaben richtig. In beiden verbalen Untertests lag der Anteil an vollständig richtig gelösten Testaufgaben zu beiden Messzeitpunkten niedriger (zwischen 0 im Untertest V2 zu Z1 und Z3 bis 14 im Untertest V2 zu Z1). Lediglich in V1 konnte ein Kind keine der Aufgaben lösen (zu Z1). Mindestens die Hälfte der Punktzahl erreichten zu beiden Messungen zwischen 77,6% (V1 zu Z3) und 88,9% (V2 zu Z3) aller Kinder.

Die Reliabilitätskennwerte sowie die Interkorrelationen der vier Subtest sind Tabelle 11.4 zu entnehmen. Die für die Berechnung der Split-Half Reliabilität erforderliche Aufteilung der Items erfolgte nach der Odd-even-Methode (r_{tt}). Aufgrund der daraus resultierenden Halbierung der Itemzahl wurden die Skalen anschließend nach Spearman-Brown um den Faktor 2 aufgewertet (Amelang & Zielinski, 2002; Bühner & M. Ziegler, 2009).

Für die beiden nonverbalen Skalen ergaben sich korrigierte Split-Half-Reliabilitäten (r_{tt_corr}) zwischen 0.80 und 0.95 für Z1 und zwischen 0.87 bis 0.93 für die zweite Messung zu Z3. Für beide verbale Skalen lagen die Reliabilitäten insbesondere zu Z1 mit $r_{tt_corr} = 0.47$ bis $r_{tt_corr} = 0.77$ deutlich niedriger. Zu Z3 ließen sich hingegen mit den nonverbalen Skalen vergleichbare Ergebnisse erzielen ($r_{tt_corr} = 0.84$ bis $r_{tt_corr} = 0.92$).

Für die Interkorrelationen der vier Subtests kann festgehalten werden, dass in beiden Messungen in jeder Kohorte die konvergenten Korrelationen innerhalb der nonverbalen bzw. der verbalen Skalen jeweils höher als die diskriminanten Korrelationen zwischen den Skalen ausfielen. Zur ersten Messung 2005 fielen die Korrelationen der nonverbalen Skalen etwas niedriger aus als die der verbalen Skalen ($r = 0.42$ bis $r = 0.58$ vs. $r = 0.54$ bis $r = 0.68$). In der zweiten Messung 2009 zeigte sich dies in K1 und K3 in gleicher Weise, allerdings fiel die Korrelation der nonverbalen Skalen in K2 höher aus als die der verbalen ($r = 0.49$ vs. $r = 0.43$). Die diskriminanten Korrelationen lagen in der Regel in moderater Höhe ($r = 0.19$ bis $r = 0.44$). Lediglich für K2 lagen die Korrelationen mit dem Subtest V2 in der zweiten Erhebung niedriger ($r = 0.15$ bzw. $r = 0.17$).

Die beobachteten positiven Interkorrelationen zwischen den vier Subtests kognitiver Fähigkeiten ließen auf das Vorliegen eines übergeordneten *g-Faktors* schließen. Zur Überprüfung dieser Annahme wurde eine Hauptkomponentenanalyse durchgeführt. Die Analyse ergab für alle Kohorten einen unrotierten ersten Faktor, der zu Z1 zwischen 50% (K2) und 55% (K1 und K3) der Varianz aufklärte. Für Z3 ergab sich eine Varianzaufklärung zwischen 47% (K2) und 55% (K1). Die entsprechenden Faktorladungen der

Tabelle 11.4
Reliabilitäten, Interkorrelationen und Faktorladungen der kognitiven Messung

		Zeitpunkt 1				Zeitpunkt 3					
		NV1	NV2	V1	V2	FL	NV1	NV2	V1	V2	FL
K_AS_2	NV1	0.82	0.51**	0.41**	0.34**	0.74		0.51**	0.29**	0.24**	0.71
	NV2		0.94	0.37**	0.28**	0.70			0.31**	0.29**	0.75
	V1			0.74	0.64**	0.82				0.47**	0.72
	V2				0.61	0.76					0.69
K1_AS_2	NV1	0.81	0.58**	0.44**	0.22**	0.77	0.93	0.46**	0.34**	0.37**	0.72
	NV2		0.95	0.35**	0.19*	0.72		0.91	0.29**	0.38**	0.71
	V1			0.67	0.58**	0.81			0.88	0.57**	0.75
	V2				0.47	0.65				0.84	0.79
K2_AS_2	NV1	0.80	0.42**	0.25**	0.30**	0.69	0.87	0.49**	0.24**	0.15	0.70
	NV2		0.93	0.28**	0.20**	0.64		0.90	0.29**	0.17	0.73
	V1			0.68	0.54**	0.75			0.84	0.43**	0.71
	V2				0.68	0.74				0.84	0.60
K3_AS_2	NV1	0.80	0.45**	0.34**	0.29**	0.68	0.92	0.56**	0.28**	0.29**	0.74
	NV2		0.93	0.35**	0.32**	0.70		0.92	0.33*	0.38**	0.80
	V1			0.77	0.63**	0.80			0.86	0.41**	0.67
	V2				0.57	0.77				0.92	0.70

Anmerkungen. N(K_AS_2) = 370-575; N(K1_AS_2) = 111-169; N(K2_AS_2) = 139-215; N(K3_AS_2) = 120-190; NV1 = Figurenklassifikation; NV2 = Figurenanalogien; V1 = Wortanalogien; V2 = Allgemeinwissen; FL = Faktorladungen; Reliabilität (Splithalf, aufgewertet nach Spearman-Brown) = Werte in der Hauptdiagonale (fett); Aufgrund der klassenspezifischen Erfassung keine Reliabilitätsangabe zu Z3 für K_AS_2.

* $p < 0.05$. ** $p < 0.01$.

Subtests können ebenfalls Tabelle 11.4 entnommen werden. Insgesamt unterstützten die hohen Ladungen auf dem jeweils extrahierten Faktor die inhaltliche Interpretation eines übergeordneten *g-Faktors*. Die Faktorwerte der Kinder wurden in der Folge als Indikator der allgemeinen kognitiven Fähigkeit genutzt, sofern keine latente Schätzung herangezogen wurde.

Die Überprüfung der *g*-Werte Verteilung ließ zudem den Schluss auf eine annähernde Normalverteilung zu. Die Verteilung erwies sich in allen Fällen als leicht linksschief (K2 zu Z3 SCH = -0.81 bis K1 zu Z1 SCH = -0.17). Zu Z1 lag eine leicht breitgipflige Verteilung in K1 und K2 vor (EX = -0.40 bzw. EX = -0.63) während für K3 eine schmalgipflige Verteilung festzustellen war (EX = 0.28). Zu Z3 war für K1 eine breitgipflige Verteilung zu beobachten (EX = -0.83), für K2 lag hingegen eine schmalgipflige Verteilung vor (EX = 1.42), bei K3 fand sich schließlich eine nahezu symmetrische Verteilung (EX = 0.03).

Zur Überprüfung der Validität des *g*-Maßes wurde der abschließend der Zusammenhang zu Schulerfolg betrachtet. Tabelle 11.5 zeigt die Korrelationen zwischen beiden Intelligenzmaßen und den fachspezifischen Schulnoten in den betrachteten Schuljahren 2004/05, 2006/07 und 2008/09. Die Zusammenhänge zwischen SE und kognitiven Fähigkeiten bei zeitgleicher Erfassung lagen jeweils am höchsten. Dabei waren die Korrelationen mit der zweiten Intelligenzmessung stärker ausgeprägt und fielen für Mathematik höher aus. Die Korrelationen für beide *g*-Maße in K1 lagen in einem mittleren bis hohen Bereich (zwischen $r = 0.25$ und $r = 0.48$). In K2 fielen die Zusammenhänge weniger deutlich aus. Hier fanden sich moderate Korrelationen zwischen dem Maß der zweiten Intelligenztestung und den Schulnoten in Mathematik (zwischen $r = 0.26$ und $r = 0.38$) für Deutsch lagen die Korrelationen durchweg in einem niedrigen Bereich (zwischen $r = 0.06$ und $r = 0.25$). Für K3 konnten hingegen mit beiden Intelligenzmaßen sowohl in Mathematik als auch in Deutsch moderate bis hohe Zusammenhänge beobachtet werden, allerdings waren auch hier diejenigen mit dem Intelligenzmaß der dritten Erhebung stärker ausgeprägt ($r = 0.16$ bis $r = 0.37$ in Z1 vs. $r = 0.35$ bis $r = 0.50$ zu Z3).

Tabelle 11.5
 Manifeste Interkorrelation zwischen dem g-Faktor und Schulnoten

		g-Faktor												
		K_AS_2			K1_AS_2			K2_AS_2			K3_AS_2			
		Z1	Z3	Z3	Z1	Z3	Z3	Z1	Z3	Z3	Z1	Z3	Z1	Z3
SE_Ma	Z1	0.21 **	0.39**	0.27*	0.25 *	0.46**	0.31**	0.30 **	0.26**	0.37**	0.20*	0.33**	0.50**	
	Z2	0.16**	0.35**	0.46**	0.45**	0.37**	0.19*	0.30**	0.26**	0.22*	0.33**	0.31**	0.33**	
	Z3	0.26**	0.35 **	0.37 **	0.40**	0.47**	0.19**	0.38**	0.06	0.33**	0.25**	0.25**	0.31**	
SE_De	Z1	0.19 **	0.33**	0.47**	0.26 *	0.41**	0.19 **	0.14	0.17	0.33 **	0.25**	0.46**	0.25**	
	Z2	0.11**	0.27**	0.41**	0.35**	0.48**	0.23**	0.25**	0.16	0.25**	0.25**	0.25**	0.25**	
	Z3	0.21**	0.37 **	0.48 **	0.36*	0.48**	0.23**	0.25 **	0.16	0.37 **	0.16	0.37 **	0.37 **	

Anmerkungen. N(K_AS_2) = 215-442; N(K1_AS_2) = 35-128; N(K2_AS_2) = 92-169; N(K3_AS_2) = 88-145; SE = Schulerfolg; Ma = Mathematik; De = Deutsch; Fett = zeitgleiche Korrelation.

* $p < 0.05$.
 ** $p < 0.01$

11.3 Motivation: Kennwerte und deskriptive Statistik

Einen Überblick über die Lage-, Streuungs- und Verteilungsmaße des fachspezifischen FSKs sowie der IM zu den drei Erhebungszeitpunkten liefert Tabelle 11.6. Wie aus dieser hervorgeht fielen die Mittelwerte des FSK und der IM über alle Kohorten und Zeitpunkte hinweg sowohl in Mathematik als auch in Deutsch mit einem durchschnittlichen Wert von etwa 4 (fünfstufige Skala mit 5 als höchstem Ausprägungsgrad) recht hoch aus (Mathematik: $M = 3.30$ bis $M = 4.14$; $SD = 0.68$ bis $SD = 1.00$; Deutsch: $M = 3.31$ bis $M = 4.00$; $SD = 0.65$ bis $SD = 0.90$). In beiden Skalen des FSKs hatten die jüngsten Kinder in der Regel zu allen Zeitpunkten die höchsten Mittelwerte (z. B. Mathematik Z2: $M = 4.14$; $SD = 0.72$ vs. $M = 3.95$; $SD = 0.72$ in K2 und $M = 3.78$; $SD = 0.86$ in K3). Lediglich zu Z3 lag der Mittelwert im FSK Mathematik für K2 etwas höher (K2: $M = 3.89$; $SD = 0.79$ vs. K1: $M = 3.83$; $SD = 0.68$). Für die IM bestätigte sich dieses Muster weitgehend. Hier lagen die Werte in der Regel ebenfalls in K1 am höchsten, lediglich zu Z3 fand sich in K2 ein etwas höherer Mittelwert im Fach Mathematik (K2: $M = 3.54$; $SD = 0.96$ vs. K1: $M = 3.44$; $SD = 0.92$). Sowohl die Mittelwerte des FSKs als auch die der IM sanken in beiden Schulfächern in den Kohorten über den Untersuchungszeitraum hinweg. Eine Ausnahme lag im FSK Deutsch in den beiden älteren Kohorten vor. Hier sanken die Mittelwerte zunächst von Z1 nach Z2 und stiegen dann nach Z3 wieder an, wobei die Ausprägung zu Z3 im Vergleich zu Z1 jeweils höher ausfiel (z. B. K3: $M = 3.73$; $SD = 0.74$ zu Z1 und $M = 3.79$; $SD = 0.74$ zu Z3).

In den Verteilungsmaßen der erhobenen Motivationsskalen zeigten sich, bedingt durch die Überrepräsentation positiver Fähigkeitsselbsteinschätzungen und einer recht hohen IM, in der Tendenz eher linksschiefe Verteilungen (die Werte der Schiefe lagen zwischen $SCH = -0.16$ in IM_Ma für K2 zu Z3 und $SCH = -1.01$ in IM_Ma für K1 zu Z1). Für den Exzess lagen Werte zwischen $EX = -0.70$ (IM_De, K1 zu Z3) und $EX = 1.41$ (FSK_De, K1 zu Z2) vor. Hier ließ sich keine klare Tendenz über die Zeitpunkte hinweg feststellen (vgl. Tabelle 11.6). Sämtliche Werte von Schiefe und Exzess lagen damit in den von West et al. (1995) angegebenen Grenzen, so dass insgesamt auch nach Inspektion der graphischen Verteilungsdiagramme in allen Kohorten und zu jedem Zeitpunkt von einer Nichtverletzung der Normalverteilungsannahme ausgegangen werden konnte.

Tabelle 11.7 informiert schließlich über die Interkorrelationen der Motivationsskalen innerhalb der Fächer und zwischen den Fächern für die drei Kohorten zu den einzelnen Zeitpunkten. Die Interkorrelationen der Skalen fielen fachkongruent jeweils am höchsten aus. Die Korrelationen für das Fach Mathematik waren dabei durchweg höher als im Fach Deutsch, wobei in beiden Fällen mittlere bis hohe Korrelationen vorlagen (Mathematik: $r = 0.65$ bis $r = 0.80$; Deutsch: $r = 0.61$ bis $r = 0.72$). Die Korrelationen in

Tabelle 11.6
Deskriptive Kennwerte und Verteilungsmaße der Motivationskalen

	Standard- abweichung									Mittelwert						Schiefe						Exzess																																
	Z 1			Z 2			Z 3			Z 1			Z 2			Z 3			Z 1			Z 2			Z 3																													
	Z 1	Z 2	Z 3	Z 1	Z 2	Z 3	Z 1	Z 2	Z 3	Z 1	Z 2	Z 3	Z 1	Z 2	Z 3	Z 1	Z 2	Z 3	Z 1	Z 2	Z 3	Z 1	Z 2	Z 3																														
FSK_Ma	K_AS_2	4.00	3.95	3.81	0.78	0.78	0.78	0.78	0.78	-0.86	-0.56	-0.45	0.68	-0.13	-0.01	K1_AS_2	4.09	4.14	3.83	0.80	0.72	0.68	-0.98	-0.75	-0.16	0.69	-0.01	-0.35	K2_AS_2	4.03	3.95	3.89	0.78	0.72	0.79	-0.88	-0.31	-0.60	0.57	-0.44	0.52	K3_AS_2	3.90	3.78	3.71	0.76	0.86	0.83	-0.78	-0.51	-0.40	1.15	-0.23	-0.42
FSK_De	K_AS_2	3.80	3.75	3.81	0.76	0.73	0.75	0.75	0.75	-0.59	-0.48	-0.64	0.02	0.17	0.28	K1_AS_2	4.00	4.00	3.93	0.71	0.68	0.65	-0.66	-0.72	-0.79	-0.06	1.41	0.12	K2_AS_2	3.72	3.69	3.74	0.79	0.68	0.82	-0.68	-0.51	-0.64	0.18	0.27	0.29	K3_AS_2	3.73	3.59	3.79	0.74	0.79	0.74	-0.36	-0.27	-0.40	-0.23	-0.33	0.00
IM_Ma	K_AS_2	3.96	3.90	3.43	0.90	0.85	0.98	0.98	0.98	-0.90	-0.59	-0.28	0.44	-0.26	-0.39	K1_AS_2	4.06	4.11	3.44	0.90	0.73	0.92	-1.01	-0.72	-0.25	0.55	0.11	-0.27	K2_AS_2	3.95	3.91	3.54	0.92	0.84	0.96	-0.95	-0.57	-0.16	0.51	-0.23	-0.62	K3_AS_2	3.88	3.69	3.30	0.88	0.92	1.00	-0.80	-0.39	-0.33	0.46	-0.63	-0.44
IM_De	K_AS_2	3.71	3.68	3.43	0.88	0.83	0.87	0.87	0.87	-0.50	-0.56	-0.36	-0.40	-0.11	-0.27	K1_AS_2	3.94	3.93	3.65	0.83	0.82	0.77	-0.85	-0.83	-0.31	0.25	-0.07	-0.70	K2_AS_2	3.63	3.63	3.31	0.89	0.83	0.90	-0.45	-0.55	-0.33	-0.32	0.24	-0.35	K3_AS_2	3.60	3.52	3.36	0.87	0.81	0.88	-0.27	-0.46	-0.30	-0.67	-0.13	-0.14

Anmerkungen. N(K_AS_2) = 320-575; N(K1_AS_2) = 112-170; N(K2_AS_2) = 145-215; N(K3_AS_2) = 135-190;
FSK = Fähigkeitsselbstkonzept; IM = Intrinsische Motivation; Ma = Mathematik; De = Deutsch.

beiden Skalen und für beide Fächer lagen zu Z2 dabei am höchsten. Sowohl die fach- als auch die konstruktübergreifenden Korrelationen (also FSK_Ma mit IM_De) fielen in der Regel am niedrigsten aus.

Die Reliabilität der Motivationsskalen wurde fachspezifisch durch Cronbachs Alpha bestimmt. Es ergaben sich zufriedenstellende bis gute interne Konsistenzen, insbesondere vor dem Hintergrund der Kürze der Skalen mit nur drei Items ($\alpha = 0.59$ für IM_De in K1 zu Z3 bis $\alpha = 0.86$ für FSK_Ma in K2 zu Z3 und K3 zu Z2). Die interne Konsistenz lag für Mathematik mehrheitlich höher als für Deutsch und für FSK meist etwas höher als für IM.

Abschließend wurden die manifesten Skaleninterkorrelationen zwischen FSK, IM und kognitiven Fähigkeiten betrachtet. Eine Analyse der Zusammenhänge zwischen den Motivationsmaßen und Schulerfolg erfolgt in Kapitel 13, da sich diese auf die Beantwortung der Fragestellungen bezieht.

Die manifesten Skaleninterkorrelationen zwischen den Motivationsskalen und dem g -Faktor sind Tabelle 11.8 zu entnehmen. Insgesamt betrachtet ergaben sich hier eher niedrige bis moderate Zusammenhänge, dabei lagen die zeitgleichen Korrelationen (fettgedruckt) in der Regel am höchsten. Außerdem zeigten sich Unterschiede zwischen den Kohorten bzw. den beiden Testungen der kognitiven Fähigkeiten. In K1 lagen die Korrelationen mit allen Motivationsmaßen zu sämtlichen Zeitpunkten in einem niedrigen Bereich (zwischen $r = -0.03$ und $r = 0.24$). Die höchsten Zusammenhänge lagen hier mit der Intelligenzmessung zu Z1 vor. Ein vergleichbares Muster fand sich auch in K2, hier lagen die Korrelationen jedoch insgesamt etwas höher (zwischen $r = 0.00$ und $r = 0.22$). Im Unterschied hierzu fanden sich in K3 moderat positive Zusammenhänge zwischen FSK, IM und beiden Intelligenzmaßen (zwischen $r = 0.00$ und $r = 0.39$).

Tabelle 11.7
Interkorrelation und Reliabilität der Motivationsskalen

	Zeitpunkt 1						Zeitpunkt 2						Zeitpunkt 3						
	Mathematik			Deutsch			Mathematik			Deutsch			Mathematik			Deutsch			
	FSK	IM	FSK	FSK	IM	FSK	FSK	IM	FSK	FSK	IM	FSK	FSK	IM	FSK	FSK	IM	FSK	FSK
K_AS_2 Mathematik	FSK	0.74	0.73 ^{**}	0.32 ^{**}	0.21 ^{**}	0.80 ^{**}	0.81	0.80 ^{**}	0.20 ^{**}	0.12 ^{**}	0.82	0.70 ^{**}	0.11 [*]	0.05					
	IM	0.73	0.19 ^{**}	0.27 ^{**}	0.27 ^{**}	0.76	0.15 ^{**}	0.20 ^{**}	0.80	0.20 ^{**}	0.11 [*]	0.03	0.22 ^{**}						
Deutsch	FSK	0.72	0.68 ^{**}	0.72	0.68 ^{**}	0.72	0.72 ^{**}	0.72 ^{**}	0.72 ^{**}	0.70	0.70	0.62 ^{**}	0.62 ^{**}						
	IM	0.71	0.71	0.71	0.71	0.71	0.70	0.70	0.70	0.70	0.67	0.67	0.67						
K1_AS_2 Mathematik	FSK	0.70	0.68 ^{**}	0.38 ^{**}	0.20 ^{**}	0.80 ^{**}	0.79	0.80 ^{**}	0.33 ^{**}	0.15	0.74	0.75 ^{**}	0.13	0.04					
	IM	0.70	0.27 ^{**}	0.35 ^{**}	0.35 ^{**}	0.68	0.27 ^{**}	0.17 [*]	0.74	0.17 [*]	0.05	0.14							
Deutsch	FSK	0.65	0.63 ^{**}	0.65	0.63 ^{**}	0.71	0.71 ^{**}	0.71 ^{**}	0.71 ^{**}	0.61 ^{**}	0.63	0.61 ^{**}							
	IM	0.64	0.64	0.64	0.64	0.75	0.75	0.75	0.75	0.59	0.59	0.59							
K2_AS_2 Mathematik	FSK	0.76	0.74 ^{**}	0.25 ^{**}	0.17 [*]	0.80 ^{**}	0.75	0.80 ^{**}	0.22 ^{**}	0.07	0.86	0.65 ^{**}	0.07	0.03					
	IM	0.75	0.14 ^{**}	0.28 ^{**}	0.28 ^{**}	0.75	0.19 ^{**}	0.22 ^{**}	0.81	0.22 ^{**}	0.06	0.21 [*]							
Deutsch	FSK	0.73	0.66 ^{**}	0.73	0.66 ^{**}	0.62	0.70 ^{**}	0.70 ^{**}	0.74	0.74 ^{**}	0.62 ^{**}	0.62 ^{**}							
	IM	0.74	0.74	0.74	0.74	0.69	0.69	0.69	0.69	0.67	0.67	0.67							
K3_AS_2 Mathematik	FSK	0.76	0.77 ^{**}	0.35 ^{**}	0.25 ^{**}	0.80 ^{**}	0.86	0.80 ^{**}	0.03	0.06	0.81	0.71 ^{**}	0.15	0.09					
	IM	0.72	0.14 [*]	0.18 ^{**}	0.18 ^{**}	0.78	-0.06	0.11	0.83	0.11	-0.03	0.28 ^{**}							
Deutsch	FSK	0.73	0.73 ^{**}	0.73	0.73 ^{**}	0.76	0.76 ^{**}	0.72 ^{**}	0.69	0.72 ^{**}	0.69	0.62 ^{**}							
	IM	0.71	0.71	0.71	0.71	0.65	0.65	0.65	0.65	0.72	0.69	0.62 ^{**}							

Anmerkungen. N(K_AS_2) = 392-574; N(K1_AS_2) = 112-170; N(K2_AS_2) = 145-215; N(K3_AS_2) = 135-190; FSK = Fähigkeitsselbstkonzept; IM = Intrinsische Motivation; Ma = Mathematik; De = Deutsch; Reliabilität (Cronbachs α) = Werte in der Hauptdiagonale (fett).

* $p < 0.05$.
 ** $p < 0.01$.

Tabelle 11.8
 Manifeste Interkorrelation zwischen *g*-Faktor und Motivation

		<i>g</i> -Faktor											
		K_AS_2			K1_AS_2			K2_AS_2			K3_AS_2		
		Z1	Z3	Z1	Z3	Z1	Z3	Z1	Z3	Z1	Z3	Z1	Z3
FSK_Ma	Z1	0.18 **	0.23**	0.17 *	0.10	0.20 **	0.22**	0.39 **	0.37**	0.39 **	0.37**	0.39 **	0.37**
	Z2	0.12**	0.06	0.20*	0.07	0.14	0.00	0.34**	0.12	0.34**	0.12	0.34**	0.12
	Z3	0.14**	0.18 **	0.04	0.03	0.15	0.13	0.30*	0.32 **	0.30*	0.32 **	0.30*	0.32 **
FSK_De	Z1	0.14 **	0.16**	0.24 *	0.14	0.22 **	0.06	0.25 **	0.33**	0.25 **	0.33**	0.25 **	0.33**
	Z2	0.03	0.13*	0.21*	0.16	0.10	0.06	0.15	0.20	0.15	0.20	0.15	0.20
	Z3	0.05	0.07	0.07	-0.04	0.11	0.00	0.09	0.28 **	0.09	0.28 **	0.09	0.28 **
IM_Ma	Z1	0.15 **	0.11*	0.11	0.03	0.20 **	0.07	0.30 **	0.25**	0.30 **	0.25**	0.30 **	0.25**
	Z2	0.07	0.02	0.14	0.08	0.13	-0.03	0.24**	0.00*	0.24**	0.00*	0.24**	0.00*
	Z3	0.11*	0.07	0.18*	-0.05	0.00	0.00	0.21*	0.22 *	0.21*	0.22 *	0.21*	0.22 *
IM_De	Z1	0.09 *	0.11**	0.13	0.12	0.18 **	0.00	0.21 **	0.26**	0.21 **	0.26**	0.21 **	0.26**
	Z2	-0.03	0.09	0.07	0.18	0.02	0.00	0.12	0.13	0.12	0.13	0.12	0.13
	Z3	0.00	0.11 **	0.03	-0.04	0.04	0.13	0.14	0.24 **	0.14	0.24 **	0.14	0.24 **

Anmerkungen. N(K_AS_2) = 370-575; N(K1_AS_2) = 111-170; N(K2_AS_2) = 139-215; N(K3_AS_2) = 120-190; FSK = Fähigkeitsselbstkonzept; IM = Intrinsische Motivation; Ma = Mathematik; De = Deutsch; Fett = zeitgleiche Korrelation.

* $p < 0.05$. ** $p < 0.01$

12 Güteprüfung der latenten Messmodelle und latente Zusammenhänge

Zur Beantwortung der unter 5.2 formulierten Fragestellungen wird auf die Modellierung von Strukturgleichungsmodellen zurückgegriffen. Die Analysen wurden mit der Software Mplus, Version 6 (L. K. Muthén & B. O. Muthén, 1998-2010) durchgeführt, wobei hier um die Abhängigkeit der Stichprobe korrigiert wurde (vgl. Kapitel 9).

12.1 Voraussetzungsprüfung

Wie unter 9.1 beschrieben, muss vor der eigentlichen Durchführung von Strukturgleichungsanalysen zunächst geprüft werden, ob die verwendeten Daten bestimmte Voraussetzungen erfüllen.

Um das Vorliegen von *Multikollinearität* auszuschließen, wurden zunächst die multiplen Korrelationen der jeweiligen Indikatoren der Prädiktoren separat berechnet, anschließend wurden sämtliche Indikatoren simultan betrachtet. Zusätzlich wurden der VIF und der Toleranz-Wert bestimmt. Grundsätzlich konnte, sowohl für Motivation als auch für die kognitiven Fähigkeiten, davon ausgegangen werden, dass zu keinem Zeitpunkt eine starke Kollinearität der Indikatoren vorlag. In allen Fällen lagen die VIF und Tolerance-Werte in den von Kline (2011) angegebenen Grenzen von $VIF > 10$ und $Tolerance < 0.10$. Dies galt sowohl in der Einzelbetrachtung als auch in der simultanen Betrachtung der Prädiktoren. Da diese Grenzen relativ liberal angesetzt sind, wurde zusätzlich der *Condition Index* (CI) bestimmt. Dieser zeigt bei Werten größer 15 eine mäßige und bei Werten größer 30 eine starke Kollinearität an (Bühner & M. Ziegler, 2009). Insbesondere für die Indikatoren der Motivation im Fach Mathematik lag der *CI* zu allen Zeitpunkten in einige Dimensionen über 15 (in keinem Fall jedoch über 30). Die beobachteten multiplen Korrelationen überschritten zwar nicht den von Tabachnick und Fidell (2001) genannten kritischen Wert von $R > 0.90$, lagen aber teilweise über $R = 0.80$, weshalb für beide Motivationsmaße im Fach Mathematik von einer mäßiger Kollinearität auszugehen ist.

Zur Identifikation *univariater Ausreißer* wurden alle Skalenwerte z-standardisiert. Über alle Skalen hinweg wichen 36 Werte mindestens drei Standardabweichungen vom Mittelwert ab ($z > |3|$). Eine Abschätzung der Relevanz dieser Ausreißer erfolgte durch einen Vergleich des Mittelwertes jeder Skala mit dem 5% getrimmten Mittel (Pallant, 2010). Dieser ergab in keinem Fall eine starke Abweichung beider Werte, von einem Ausschluss dieser Fälle wurde daher abgesehen.

Zur Identifikation *multivariater Ausreißer* wurden für jede Kohorte Mahalanobis Distanzen berechnet (Tabachnick & Fidell, 2001) und mit dem kritischen χ^2 -Wert ($\chi^2(20) = 45.31, p < 0.00$) verglichen. In jeder Kohorte konnte so je 1 Fall als multivariater Ausreißer identifiziert werden. Diese wurden in der Folge aus den Strukturgleichungsanalysen ausgeschlossen.

Zur Durchführung von Strukturgleichungsanalysen werden vollständige Datensets benötigt. Fehlende Werte können jedoch bei Vorliegen eines unsystematischen Ausfallmusters adäquat ersetzt werden (vgl. 9.1). Dementsprechend wurden die fehlenden Werte in der Gesamtstichprobe für jeden Zeitpunkt sowie über den gesamten Studienverlauf hinweg auf einen systematischen Ausfall mit dem MCAR Test nach Little (R. J. A. Little, 1988) geprüft. Hierzu wurde die *Missing Value Analyse* (MVA) des Statistikpaketes SPSS, Version 18 genutzt. Die Verteilung der fehlenden Werte wurde auf die Bedingungen *MCAR* und *MAR* geprüft (Hill, 1997). Insgesamt zeigten sich für die Motivationsvariablen zu allen Zeitpunkten sehr geringe Missing-Raten. Diese lagen für den 1. Erhebungszeitpunkt zwischen 0 und 1,7%, für den 2. Messzeitpunkt bei bis zu 1% und für die letzte Messung zwischen 0,8 und 2,8%. Für die kognitiven Fähigkeiten fanden sich auf Itemebene durch die Konzeption der Testbatterie als Power-Speed Test (vgl. Abschnitt 8.3.2) erwartungsgemäß höhere Missingraten von bis zu 23%. In den Noten fanden sich vor allem für das Schuljahr 2004/05 relativ viele fehlende Werte (zwischen 26% und 44%), was dem Umstand geschuldet ist, dass in vielen Grundschulen erst am Ende der 2. Klasse Zeugnisnoten vergeben werden. Für die erste Erhebung ergab eine MVA aller Variablen, dass die MCAR Annahme nicht gehalten werden konnte ($\chi^2(587, N = 560) = 882.89, p = 0.00$), nach Inspektion der Ergebnisse des im Rahmen der MVA ausgegebenen t-Tests konnte jedoch zumindest von einem *at random* Ausfall ausgegangen werden. Zu beiden anderen Zeitpunkten bestätigte der MCAR Test eine *completely at random* Verteilung der Missings in allen Variablen: zu Z2 $\chi^2(83, N = 324) = 89.27, p = 0.30$ und zu Z3 $\chi^2(313, N = 370) = 325.61, p = 0.30$. In der Analyse der Indikatoren über die Zeit hinweg konnte für das FSK, die IM, die kognitiven Fähigkeiten und schließlich für die Noten im Fach Mathematik von der Haltbarkeit der MCAR Annahme ausgegangen werden ($\chi^2(309, N = 389) = 358.47, p = 0.03$, $\chi^2(242, N = 381) = 272.98, p = 0.08$, $\chi^2(8, N = 370) = 18.58, p = 0.02$ und $\chi^2(53, N = 190) = 71.07, p = 0.05$). Für die Schulnoten im Fach Deutsch fiel der MCAR Test signifikant aus ($\chi^2(56, N = 187) = 88.80, p = 0.00$), die MAR Annahme

bestätigte sich jedoch. Für die durchzuführenden Analysen wurden in der Folge Kinder ausgewählt, die an mindestens zwei der drei Messzeitpunkte teilgenommen haben. Für die Folgeanalysen wurden die fehlenden Daten mit dem unter 9 beschriebenen *EM-Algorithmus* ersetzt.

Als weitere Voraussetzung wurde das Vorliegen *multivariater Normalverteilung*¹ betrachtet. Eine Prüfung der multivariaten Normalverteilung erlaubt der Mardia-Test. Dieser zeigte an, dass in der vorliegenden Stichprobe K_AS_2 zumindest für einige Messmodelle von der Verletzung dieser Annahme ausgegangen werden muss (Multivariate Kurtosis = 1.51, *c.r.* = 2.5 bis 10.45, *c.r.* = 17.22). Dennoch wies keines der Items einen Wert der Schiefe und Kurtosis auf, der nicht innerhalb der von West et al. (1995) geforderten Grenzen lag. Nach Bühner (2006) kann in einem solchen Fall trotz Verletzung der Multinormalverteilungsannahme bei Vorliegen einer Stichprobe > 100 auf eine ML-Schätzung zur Strukturgleichungsanalyse zurückgegriffen werden, da sich dieses Verfahren als „ausgesprochen robust gegenüber Verletzungen der Normalverteilungsannahme“ (Bühner, 2006, S.251) erwiesen hat. Von dieser sollte nur im Fall einer extremen Verletzung der Multinormalverteilungsannahme, die hier nicht zu konstatieren ist, Abstand genommen werden (Bollen, 1989).

Die Interpretation der χ^2 -Werte sollte dennoch unter Berücksichtigung der Verletzung Verteilungsannahme vorgenommen werden. Diese können nach B. M. Byrne (2001) bei nicht multinormalverteilten Daten überhöht sein, was bedeutet, dass ein eigentlich passendes Modell zu häufig abgelehnt wird. Aus diesem Grund werden nachfolgend in erster Linie die Fit-Indikatoren zur Interpretation der Strukturgleichungsmodelle herangezogen, die χ^2 -Werte der Vollständigkeit halber jedoch ebenfalls berichtet. Des Weiteren ist bei der Interpretation der Parameterschätzungen zu beachten, dass diese aufgrund der Unterschätzung des Standardfehlers unter Umständen Signifikanz erreichen, obwohl dies in der Population nicht zutreffen würde (Bühner & M. Ziegler, 2009). Da gleichzeitig durch die Berücksichtigung der Clusterung der Zwillingsdaten in Mplus der Standardfehler höher ausfallen wird (siehe dazu Kapitel 9) wird bei der Interpretation der Ergebnisse die numerische Höhe der Parameter im Mittelpunkt stehen, da diese unabhängig vom Standardfehler sind.

Als weitere Voraussetzung wurde die *Linearität und Homoskedastizität* der Daten geprüft. Durch Inspektion der bivariaten Streudiagramme mit zusätzlich geplotteter Regressionsgerade (95% Konfidenzintervall) konnte diese als erfüllt angesehen werden.

Zur Überprüfung der *Varianzhomogenität* zwischen den drei Kohorten wurde der Levene-Test herangezogen. Die Varianzen stellten sich zu Z1 für 9, zum Z2 für 4

¹ Eine Prüfung der univariaten Normalverteilung erfolgte bereits in Kapitel 11.

Indikatoren und zum letzten Zeitpunkt für 7 Indikatoren als signifikant unterschiedlich heraus. Tabachnick und Fidell (2001) empfehlen bei inhomogenen Varianzen die Durchführung eines F_{max} -Tests, um das Ausmaß der Varianzinhomogenität abschätzen zu können. Sämtliche Ergebnisse des F_{max} -Tests ergaben Werte zwischen 1.08 und 1.79. Die beobachteten Varianzunterschiede zwischen den Gruppen konnten demnach als vernachlässigbar angesehen werden.

12.2 Modell-Fit der Messmodelle und Invarianzprüfung

Vor der Überprüfung der Vorhersage schulischen Erfolgs durch FSK und IM unter Berücksichtigung der allgemeinen kognitiven Fähigkeit im MKSD wurden die Messmodelle der Variablen hinsichtlich ihrer Passung geprüft. Den Annahmen von Marsh und Hau (1996) folgend wurden Modelle mit korrelierten Messfehlern über die Zeit (CU) gegen Modelle ohne CU getestet. Ohne Modellierung der CU wird die Stabilität der latenten Variablen in der Regel über- und die später zu spezifizierenden CL-Pfade unterschätzt. Hieran anschließend wurde die Invarianz der Messmodelle über die Zeit geprüft um sicherzustellen, dass in den Kohorten zu den unterschiedlichen Zeitpunkten auch Gleiches erfasst wurde.

12.2.1 Messmodell des schulischen Erfolgs

Zur Operationalisierung schulischen Erfolgs wurde getrennt für Mathematik und Deutsch ein Messmodell über die drei Erhebungszeitpunkte spezifiziert (vgl. Anhang A). Als Indikatoren wurde jeweils die Halbjahres- und Jahresendnote im entsprechenden Schuljahr herangezogen. Tabelle 12.1 sind die Fit-Indikatoren getrennt für jede Kohorte zu entnehmen, die Faktorladungen der Indikatoren des bestfittenden Modells werden in Tabelle B.4 (siehe Anhang B) aufgeführt. Es wurde jeweils ein Modell mit unkorrelierten Fehlern gegen ein Modell mit korrelierten Fehlern über die Zeit getestet. Aus Gründen der Identifizierbarkeit des Messmodells konnten Fehlerkorrelationen lediglich zwischen den Noten benachbarter Messzeitpunkte zugelassen werden.

Wie Tabelle 12.1 zu entnehmen ist, wiesen die Modelle mit korrelierten Fehlern ($M1$) zwischen den Zeitpunkten bis auf eine Ausnahme einen besseren Modell-Fit auf, als Modelle ohne korrelierte Fehler. Lediglich in K1 für SE Mathematik führte die Schätzung der *correlated uniqueness* zu keiner Verbesserung im Modell-Fit. In diesem Fall lagen jedoch nicht positiv definite (siehe zu diesem Aspekt Kline, 2011) Matrizen vor, weshalb auch für K1 Modell $M1$ als Messmodell angenommen wurde. Mit diesen wurden in der Folge alle weiteren Analysen durchgeführt. Die Messmodelle für beide Schulfächer

lieferten in allen Kohorten einen exzellenten Fit (vgl. Tabelle 12.1), dieser war im Fach Deutsch insgesamt etwas höher als in Mathematik. In Mathematik lag der Fit in K2 dabei am höchsten ($\chi^2 = 1.47$, $df = 2$, $p = 0.48$; CFI = 1.00; TLI = 1.01; SRMR = 0.01). Hinsichtlich der Faktorladungen wiesen beide Noten in allen Schuljahren insgesamt sehr hohe Ladungen auf, die immer in etwa gleicher Höhe lagen (standardisierte Lösung: $\lambda = 0.79$ bis $\lambda = 0.99$). Für Deutsch ergaben sich in allen Kohorten für die Jahresendnote im Schuljahr 2004/05 nicht signifikante negative Fehlervarianzen von -0.01 , was einen Hinweis auf eine Fehlspezifikation des Modells darstellt. Allerdings wurden Ausreißer vor der Modellschätzung eliminiert und können daher keine Ursache dieser negative Fehlervarianz darstellen (Kline, 2011). Da die Fehlervarianzen sehr niedrig ausfielen und die Modellschätzung konvergierte, wurden die Fehlervarianzen in der Folge für K2 und K3 auf 0.01 fixiert (Chen, Bollen, Paxton, Curran & Kirby, 2001). In K1 fand sich hier der beste Modell-Fit ($\chi^2 = 1.27$, $df = 2$, $p = 0.74$; CFI = 1.00; TLI = 1.02; SRMR = 0.01). Auch für Deutsch lagen alle Faktorladungen in einem hohen Bereich, in K2 und K3 stellte die Halbjahresnote jeweils den leicht besseren Indikator dar (standardisierte Lösung: $\lambda = 0.84$ bis $\lambda = 0.97$).

12.2.2 Messmodell der kognitiven Fähigkeiten

Der Modell-Fit des Messmodells der kognitiven Fähigkeiten ist in Tabelle 12.2 dargestellt. Es wurde zunächst ein Modell mit zwei latenten g -Faktoren mit den jeweils vier Subtests als Indikatoren geschätzt ($M2$). Zwischen beiden Faktoren wurde eine Korrelation zugelassen. Wie aus Tabelle 12.2 hervorgeht, lieferte dieses Modell eine sehr schlechte Anpassung an die empirischen Daten, insbesondere in K2 ($\chi^2 = 98.34$, $df = 15$, $p = 0.00$; CFI = 0.62; TLI = 0.30; SRMR = 0.08). Wurden zusätzlich Fehlerkorrelationen innerhalb eines Zeitpunktes zwischen den verbalen und nonverbalen Subtests spezifiziert ($M1$), verbesserte sich der Modell-Fit zwar etwas, blieb aber insgesamt in einem nicht akzeptablen Bereich, wiederum vor allem in K2 ($\chi^2 = 32.72$, $df = 11$, $p = 0.00$; CFI = 0.89; TLI = 0.75; SRMR = 0.05). Aus diesem Grund wurden in einem anschließenden Schritt die Modelle für beide latente Faktoren separat geschätzt ($M3$ und $M4$). Es zeigte sich, dass der schlechte Modell-Fit des Gesamtmodells auf die schlechte Passung des ersten g -Faktors in K2 zurückzuführen war ($\chi^2 = 4.07$, $df = 1$, $p = 0.00$; CFI = 0.97; TLI = 0.84; SRMR = 0.02). Für den zweiten g -Faktor war der Fit eines Modells mit korrelierten Fehlern zwischen den verbalen und nonverbalen Subtests hingegen für alle Kohorten sehr gut (vgl. Anhang A). Aufgrund der Ergebnisse der latenten Modellierung und der unter Abschnitt 11.2 berichteten niedrigen Zusammenhänge des manifesten g -Faktors aus $Z1$ mit SE wurde die erste Intelligenzmessung aus allen weiteren Analysen ausgeschlossen. Dies hatte jedoch zur Konsequenz, dass bei längsschnittlichen Modellen, in denen die kognitiven Fähigkeiten berücksichtigt werden

Tabelle 12.1
Modell-Fit des Messmodells – Schulerfolg Mathematik und Deutsch

	Modell	Vergleich	χ^2	df	χ^2/df	CFI	TLI	SRMR	AIC	Δdf	$\Delta \chi^2 (TRd)$
SE_Ma	K1_AS_2	M1	7.21*	2	3.61	0.99	0.92	0.01	1432.72		
	M2	M1	7.51	6	1.25	0.99	0.99	0.01	1428.98	4	0.21
	K2_AS_2	M1	1.47	2	0.74	1.00	1.01	0.01	2286.63		
	M2	M1	17.59**	6	2.93	0.98	0.95	0.01	2296.43	4	13.74**
	K3_AS_2	M1	10.74**	2	5.37	0.98	0.85	0.01	2144.78		
	M2	M1	17.75**	6	2.96	0.98	0.93	0.02	2146.87	4	9.54*
SE_De	K1_AS_2	M1	1.27	2	0.64	1.00	1.01	0.00	1271.94		
	M2	M1	8.51	6	1.42	0.99	0.99	0.01	1270.6	4	9.92*
	M3	M1	1.27	3	0.42	1.00	1.02	0.01	1270.55	1	0.00
	K2_AS_2	M1	3.86	2	1.93	0.99	0.98	0.01	1893.15		
	M2	M1	9.26	6	1.54	0.99	0.99	0.01	1891.02	4	5.22
	M3	M1	3.63	3	1.21	1.00	1.00	0.01	1891.95	1	-0.13
K3_AS_2	M1		2.08	2	1.04	1.00	0.99	0.01	1751.05		
	M2	M1	7.41	6	1.24	0.99	0.99	0.01	1748.14	4	5.63
	M3	M1	2.46	3	0.82	1.00	1.01	0.01	1749.21	1	0.44

Anmerkungen. N(K1_AS_2) = 170; N(K2_AS_2) = 215; N(K3_AS_2) = 190; SE = Schulerfolg; Ma = Mathematik; De = Deutsch; CFI = Comparative Fit Index; TLI = Tucker-Lewis-Index; SRMR = Standardized Root Mean Square Residual; AIC = Akaike Information Criterion; Δdf = Differenz der Freiheitsgrade genesteter Modelle; $\Delta \chi^2 (TRd)$ = Satorra-Bentler Chi-Quadrat Differenztest; M1 = Modell mit korrelierten Fehlern (CU); M2 = Modell ohne korrelierte Fehler; M3 = Modell mit fixierter Fehlervarianz der Jahresendnote 04/05; Die letztlich ausgewählten Modelle der weiteren Analysen sind *fett* markiert.

* $p < 0.05$. ** $p < 0.01$.

sollten, nun nur noch eine Intelligenzmessung zur Verfügung stand und es sich damit für $Z1$ und $Z2$ nicht mehr um eine zeitgleiche Erfassung der Prädiktoren bzw. des Kriteriums handelte. Unter Berücksichtigung der unter 2.4 zusammengefassten empirischen Studien zur Stabilität von Intelligenz, wird dieses Vorgehen jedoch als legitim erachtet. Bei der Interpretation der Ergebnisse wird diesem Umstand entsprechend Rechnung getragen.

Die Faktorladungen der vier Subtests lagen in K1 sowie K3 in einem moderaten Bereich und fielen dabei für die verbalen und nonverbalen Subtests vergleichbar hoch aus (standardisierte Lösung: $\lambda = 0.47$ bis $\lambda = 0.62$). In K2 lagen die Faktorladungen mit Ausnahme von $V1$ niedriger (standardisierte Lösung: $\lambda = 0.34$ bis $\lambda = 0.69$).

12.2.3 Messmodell von FSK und IM

Der Fit der Messmodelle für FSK sowie IM kann Tabelle 12.3 bzw. 12.4 entnommen werden. Für FSK und IM wurde fachspezifisch ein Modell mit drei Indikatoren je Zeitpunkt spezifiziert (vgl. Anhang A). Wie bereits für SE lieferte ein Modell mit korrelierten Messfehlern über die Zeit immer die bessere Anpassung an die Daten.

Für FSK fiel der Fit für die Messmodelle in Mathematik besser aus als derjenige in Deutsch und zeigte für die jüngsten Kindern immer die beste Passung. Insgesamt ist der Fit des Messmodells FSK Mathematik als sehr gut zu bewerten (in K1: $\chi^2 = 9.21$, $df = 15$, $p = 0.87$; CFI = 1.00; TLI = 1.04; SRMR = 0.03), aber auch in Deutsch lag zumindest ein noch akzeptabler Modell-Fit vor (insbesondere K3: $\chi^2 = 45.59$, $df = 15$, $p = 0.00$; CFI = 0.94; TLI = 0.86; SRMR = 0.05). Mit Blick auf die Faktorladungen (vgl. Anhang B) als Hinweis auf die Reliabilität der Indikatoren erwies sich insbesondere Item 1 („Wie gut bist Du im Lösen von Zahlenrätseln und Textaufgaben?“) zu den beiden ersten Messzeitpunkten mit standardisierten Faktorladungen zwischen $\lambda = 0.41$ bis $\lambda = 0.71$ als schwächster Indikator. Dies war vor allem in den jüngeren Kohorten zu beobachten. In Deutsch lagen die Faktorladungen meist in einem moderaten Bereich (standardisierte Lösung: $\lambda = 0.42$ bis $\lambda = 0.91$).

Bei den Messmodellen der IM fand sich ebenfalls für Mathematik eine generell bessere Passung. Für das Messmodell in Mathematik lagen in K1 und K2 negative Fehlerkorrelationen mit Item 3 vor, weshalb diese aus dem Messmodell entfernt wurde ($M3$). Die Modelltestung dieses Modells gegen das vollständige Modell mit korrelierten Fehlern aller Items ($M1$) erbrachte für K1 und K2 keine signifikant schlechtere Passung, wohingegen in K3 ein Modell ohne Fehlerkorrelation von Item 3 schlechter fittete als $M1$. Dennoch konnte auch in K3 ein guter Fit von $M3$ erzielt werden. Um Äquivalenz der Messmodelle zwischen den Kohorten herzustellen wurde in K3 ebenfalls $M3$ als Messmodell angenommen. Die beste Passung konnte wiederum in K1 erzielt werden

Tabelle 12.2
Modell-Fit des Messmodells – Kognitive Fähigkeiten

	Modell	Vergleich	χ^2	df	χ^2/df	CFI	TLI	SRMR	AIC	Δdf	$\Delta\chi^2(TRd)$
K1_AS_2	M1		25.25 ^{**}	11	2.30	0.95	0.88	0.05	6209.34		
	M2	M1	67.93 ^{**}	15	4.53	0.83	0.67	0.08	6247.77	4	29.13 ^{**}
	M3		0.01	1	0.01	1.00	1.03	0.00	3840.75		
	M4		1.19	1	1.19	1.00	0.99	0.02	2401.84		
K2_AS_2	M1		32.72 ^{**}	11	2.97	0.89	0.75	0.05	7462.08		
	M2	M1	98.34 ^{**}	15	6.56	0.62	0.30	0.08	7502.80	4	69.40 ^{**}
	M3		4.07 [*]	1	4.07	0.97	0.84	0.02	4619.33		
	M4		0.05	1	0.05	1.00	1.09	0.00	2858.57		
K3_AS_2	M1		19.52 [*]	11	1.77	0.97	0.92	0.05	6534.98		
	M2	M1	46.94 ^{**}	15	3.13	0.89	0.79	0.09	6559.48	4	19.37 ^{**}
	M3		0.10	1	0.10	1.00	1.04	0.00	4024.92		
	M4		0.14	1	0.14	1.00	1.07	0.01	2561.44		

Anmerkungen. N(K1_AS_2) = 170; N(K2_AS_2) = 215; N(K3_AS_2) = 190; CFI = Comparative Fit Index; TLI = Tucker-Lewis-Index; SRMR = Standardized Root Mean Square Residual; AIC = Akaike Information Criterion; Δdf = Differenz der Freiheitsgrade genesteter Modelle; $\Delta\chi^2(TRd)$ = Satorra-Bentler Chi-Quadrat Differenztest; M1 = Modell mit 2 korrelierten latenten Faktoren, korrelierten Messfehlern über die Zeit (CU) sowie korrelierten Fehlern zwischen V1 und V2 sowie NV1 und NV2; M2 = Modell mit 2 korrelierten latenten Faktoren (Z1 und Z3) und korrelierten Messfehlern über die Zeit (CU); M3 = latenter *g*-Faktor Z1 mit korrelierten Fehlern zwischen V1 und V2 sowie NV1 und NV2; M4 = latenter *g*-Faktor Z3 mit korrelierten Fehlern zwischen V1 und V2 sowie NV1 und NV2; Die letztlich ausgewählten Modelle der weiteren Analysen sind *fett* markiert.

* $p < 0.05$. ** $p < 0.01$.

($\chi^2 = 7.78$, $df = 18$, $p = 0.98$; CFI = 1.00; TLI = 1.08; SRMR = 0.03). Wie schon beim FSK Mathematik beobachtet, zeigte sich auch bei der IM in Mathematik, dass Item 1 („Wie sehr magst Du das Lösen von Zahlenrätseln und Textaufgaben?“) zu allen Messzeitpunkten meist nur mäßige Faktorladungen auf den latenten IM Faktor aufwies (zwischen $\lambda = 0.41$ bis $\lambda = 0.57$; vgl. Anhang B), wohingegen die verbleibenden Items sehr hoch auf den Faktor luden (zwischen $\lambda = 0.63$ bis $\lambda = 0.99$).

Für die IM Deutsch lag für Item 1 („Wie sehr magst Du schreiben (z. B. Aufsätze)“?) zum dritten Messzeitpunkt eine hohe negative Fehlervarianz (bis zu -0.13) vor, weshalb dieses nicht als Indikator von IM zu Z3 verwendet wurde (M4). Der Fit des resultierenden Modells war, insbesondere für K3 in einem hohen Bereich (K3: $\chi^2 = 8.32$, $df = 10$, $p = 0.59$; CFI = 1.00; TLI = 1.01; SRMR = 0.03).

12.2.4 Überprüfung der Messinvarianz über die Zeit

Zur Überprüfung der Messäquivalenz der latenten Faktoren von FSK, IM und SE über die Zeit wurden die unter Kapitel 9.3 beschriebenen Stufen der Messinvarianz geprüft. Um von Messkonstanz ausgehen zu können, sollte mindestens die Bedingung der *schwachen faktoriellen Invarianz* der Faktoren erfüllt sein (Weiber & Mühlhaus, 2010). *Konfigurale Invarianz* kann hier insgesamt angenommen werden, da sämtliche Messmodelle zumindest einen akzeptablen Modell-Fit aufwiesen. Die Modellprüfung ergab insgesamt, dass die Annahme der Messinvarianz als gegeben betrachtet werden konnte. Die resultierenden Modellschätzungen finden sich in Anhang B.

Genauer konnte für die latente Modellierung des schulischen Erfolgs in Mathematik (siehe Tabelle B.5) für K1 schwache faktorielle Invarianz (M2) zunächst nicht bestätigt werden. Ein Freisetzen der Ladung der Jahresendnote 2004/05 (M4) führte jedoch zu keiner signifikanten Verschlechterung des Modell-Fits mehr, es konnte daher in der Folge zumindest von einer partiellen Messinvarianz ausgegangen werden (Geiser, 2010). In K2 konnte sowohl eine schwache (M2) als auch eine starke (M3) faktorielle Invarianz bestätigt werden. In K3 lag ebenfalls schwache faktorielle Invarianz vor. Für die latente Modellierung des schulischen Erfolgs in Deutsch konnte in allen Kohorten von starker faktorieller Invarianz ausgegangen werden (siehe Tabelle B.5).

Für FSK Mathematik (siehe Tabelle B.6) lag in K2 und K3 starke faktorielle Invarianz vor (M3), wohingegen in K2 lediglich von partieller Invarianz ausgegangen werden konnte (M4; Ladung von Item 2 frei geschätzt). Sowohl im FSK Deutsch als auch in der IM Mathematik sowie IM Deutsch für K1 und K3 konnte eine schwache faktorielle Invarianz bestätigt werden. Hingegen musste für K2 in IM Deutsch die Ladung von Item 2 wieder freigesetzt werden, um zumindest partielle faktorielle Invarianz annehmen zu können (siehe Tabelle B.7).

Tabelle 12.3
Modell-Fit des Messmodells – FSK Mathematik und Deutsch

	Modell	Vergleich	χ^2	df	χ^2/df	CFI	TLI	SRMR	AIC	Δdf	$\Delta\chi^2(TRd)$
FSK_Ma	K1_AS_2	M1	9.21	15	0.61	1.00	1.04	0.03	3052.77		
	M2	M1	31.48	24	1.31	0.98	0.96	0.05	3059.98	9	17.33*
	K2_AS_2	M1	27.39*	15	1.83	0.98	0.94	0.05	3773.37		
	M2	M1	56.07**	24	2.34	0.94	0.9	0.05	3789.38	9	23.20**
	K3_AS_2	M1	20.28	15	1.35	0.99	0.97	0.05	3338.82		
	M2	M1	46.92**	24	1.96	0.95	0.93	0.06	3356.92	9	18.56*
FSK_De	K1_AS_2	M1	13.29	15	0.89	1.00	1.02	0.04	3015.69		
	M2	M1	55.85**	24	2.33	0.86	0.79	0.07	3045.51	9	40.25**
	K2_AS_2	M1	26.76*	15	1.78	0.97	0.94	0.05	4004.42		
	M2	M1	196.59**	24	8.19	0.63	0.44	0.08	4129.67	9	272.02**
	K3_AS_2	M1	45.59**	15	3.04	0.94	0.86	0.05	3396.11		
	M2	M1	142.84**	24	5.95	0.77	0.65	0.08	3482.07	9	80.24**

Anmerkungen. N(K1_AS_2) = 170; N(K2_AS_2) = 215; N(K3_AS_2) = 190; FSK = Fähigkeitsselbstkonzept; Ma = Mathematik; De = Deutsch; CFI = Comparative Fit Index; TLI = Tucker-Lewis-Index; SRMR = Standardized Root Mean Square Residual; AIC = Akaike Information Criterion; Δdf = Differenz der Freiheitsgrade genesteter Modelle; $\Delta\chi^2(Trd)$ = Satorra-Bentler Chi-Quadrat Differenztest; M1 = Modell mit korrelierten Fehlern (CU); M2 = Modell ohne korrelierte Fehler; Die letztlich ausgewählten Modelle der weiteren Analysen sind *fett* markiert.

* $p < 0.05$. ** $p < 0.01$.

Tabelle 12.4
Gütemodelle – Motivation – IM Mathematik und Deutsch

	Modell	Vergleich	χ^2	df	χ^2/df	CFI	TLI	SRMR	AIC	Δdf	$\Delta\chi^2(\text{TRd})$
IM_Ma	K1_AS_2	M1	5.45	15	0.36	1.00	1.09	0.30	3475.82		
	M2	M1	29.34	24	1.22	0.98	0.97	0.05	3483.30	9	19.54*
	M3	M1	7.78	18	0.43	1.00	1.08	0.03	3478.49	3	1.78
K2_AS_2	M1		10.84	15	0.72	1.00	1.02	0.03	4405.41		
	M2	M1	33.52	24	1.40	0.98	0.97	0.05	4414.50	9	18.22*
	M3	M1	12.69	18	0.71	1.00	1.00	0.03	4402.56	3	1.14
K3_AS_2	M1		19.76	15	1.32	0.99	0.98	0.05	3897.50		
	M2	M1	57.73**	24	2.41	0.93	0.89	0.08	3926.10	9	30.48**
	M3	M1	28.92*	18	1.61	0.98	0.95	0.06	3901.59	3	9.06*
IM_De	K1_AS_2	M4	23.66**	10	2.37	0.94	0.83	0.06	3118.64		
	M5	M4	105.88**	17	6.23	0.92	0.37	0.11	3178.76	7	98.84**
K2_AS_2	M4		17.70	10	1.77	0.98	0.94	0.05	4127.37		
	M5	M4	163.93**	17	9.64	0.61	0.36	0.11	4251.45	7	158.90**
K3_AS_2	M4		8.32	10	0.83	1.00	1.01	0.03	3518.70		
	M5	M4	130.08**	17	7.65	0.67	0.45	0.12	3636.93	7	94.79**

Anmerkungen. N(K1_AS_2)=170; N(K2_AS_2)=215; N(K3_AS_2)=190; IM=Intrinsische Motivation; Ma=Mathematik; De=Deutsch; CFI=Comparative Fit Index; TLI=Tucker-Lewis-Index; SRMR=Standardized Root Mean Square Residual; AIC=Akaike Information Criterion; Δdf =Differenz der Freiheitsgrade generierter Modelle; $\Delta\chi^2(\text{TRd})$ =Satorra-Bentler Chi-Quadrat Differenztest; M1=Modell mit korrelierten Fehlern (CU); M2=Modell ohne korrelierte Fehler; M3=Modell ohne Fehlerkorrelation von Item 3; M4=Modell ohne Item 2 als Indikator zu Z3, mit korrelierten Fehlern; M5=M4 ohne korrelierte Fehler; Die letztlich ausgewählten Modelle der weiteren Analysen sind *fett* markiert.

* $p < 0.05$. ** $p < 0.01$.

Die Ergebnisse der Voraussetzungsprüfung sowie der latenten Modellierung der Messmodelle ließen darauf schließen, dass die Güte der resultierenden Modelle hinreichend gut war, um in der Folge eine latente Modellierung für die Prüfung der phänotypischen Fragestellungen zugrunde zu legen. Auch die Überprüfung der Messinvarianz der Messmodelle über die Zeit belegte die Angemessenheit der Annahme, dass die Konstrukte zu allen Zeitpunkten das Gleiche erfassen. Diese notwendige Voraussetzung zur Modellierung im längsschnittlichen Design konnte demnach als erfüllt betrachtet werden.

Im Folgenden Kapitel werden die Ergebnisse zur Überprüfung der phänotypischen Fragestellungen, basierend auf den zuvor bestätigten Messmodellen der betrachteten Konstrukte, dargestellt. Hieran schließt sich die Beantwortung der ätiologischen Fragestellungen an.

13 Ergebnisse zur Beantwortung der phänotypischen Fragestellungen

13.1 Intra- und Interindividuelle Stabilität

Die Beantwortung der unter Kapitel 5.1 formulierten Hypothesen bezüglich der intra- und interindividuellen Stabilität der Variablen wurde unter Rückgriff auf Strukturgleichungsmodellierungen vorgenommen. Ausgehend von den in 12.2 identifizierten Best-Fit Modellen wurden jeweils Modelle mit spezifischen Restriktionen gegen das vollständige Modell auf signifikante Unterschiedlichkeit geprüft. Durch den Wegfall der kognitiven Fähigkeitserfassung zu $Z1$ konnten die bezüglich der intra- und interindividuellen Stabilität formulierten Hypothesen zu allgemeinen kognitiven Fähigkeiten nicht geprüft werden.

13.1.1 Intraindividuelle Stabilität

In einem ersten Schritt wurden die Mittelwerte der latenten Variablen von FSK, IM und SE zu $Z1$, $Z2$ und $Z3$ gleichgesetzt und geprüft, ob hieraus eine signifikante Verschlechterung des Modell-Fits im Vergleich zum vollständigen Modell resultierte. Ist dies der Fall, kann von einer bedeutsamen Veränderung in den Mittelwerten der Variablen über den Untersuchungszeitraum hinweg ausgegangen werden. In der Folge wäre dann zu prüfen, ob sich vor allem am Übergang von der Grundschule zur Sekundarschule eine Veränderung zeigt. Die Überprüfung der Veränderung der Varianz erfolgte hierzu analog.

Die Ergebnisse der Modellprüfung der Variablen SE, FSK und IM in den beiden Fächern Mathematik und Deutsch sind den Tabellen 13.1, 13.2 sowie 13.3 zu entnehmen. Wie hieraus hervorgeht ergaben sich für einige Submodelle negative χ^2 Differenzen, was aus der Korrektur um genestete Daten resultieren kann (L. K. Muthén & C. Weber, 2010). Diese Ergebnisse dürfen in der Folge nicht interpretiert werden. Die resultierenden latenten Mittelwerte und Varianzen sind in Tabelle 13.4 enthalten.

Werden die Schulnoten in Mathematik und Deutsch betrachtet kann zunächst festgehal-

ten werden, dass diese hoch ausfielen und dabei in beiden Fächern ungefähr vergleichbar waren. Mit Blick auf den Modellvergleich (vgl. Tabelle 13.1) kann jedoch konstatiert werden, dass in keinem Fall eine signifikante Verschlechterung des Modell-Fits durch Gleichsetzung der Mittelwerte über die Zeit resultierte. Bezüglich einer Varianzveränderung konnte, mit Ausnahme des schulischen Erfolgs in Deutsch für K1 und K2¹, eine signifikante Veränderung der Varianz festgestellt werden. Die Modelle mit gleichgesetzten Varianzen (*M3*) zeigten eine signifikant schlechtere Passung als die vollständigen Modelle (*M1*). Während sich in Mathematik die Varianz mit der Zeit vergrößerte, zeigte sich für K2 und K3 in Deutsch ein umgekehrter Trend. Für die Schulnoten in Mathematik in K1 erhöhte sich die Varianz direkt nach dem Übergang (von *Z2* nach *Z3*) von 0.37 auf 0.75. In den beiden älteren Kohorten zeigte sich hingegen zunächst weniger Varianz in den Noten nach dem Übergang in die weiterführende Schule (von *Z1* nach *Z2*) und erst danach ein deutlicherer Anstieg in den darauf folgenden Schuljahren. In Deutsch lagen die Varianzen zu *Z1* zwischen 0.53 bis 0.57 und zu *Z3* zwischen 0.35 und 0.44. Die Varianz der Schulnoten in Mathematik und Deutsch lag damit zu sämtlichen Messzeitpunkten in allen Kohorten in einem eher niedrigen Bereich (von 0.33 bis 0.75; vgl. Tabelle 13.4).

Wie aus Tabelle 13.2 und 13.3 weiterhin hervorgeht, fanden sich bezüglich der Motivationsskalen vor allem in der IM bedeutsame Mittelwertsveränderungen (*M2*), während dies im FSK lediglich in K1 für Mathematik feststellbar war. In keinem Fall fittete ein Modell mit gleichgesetzten Varianzen (*M3*) signifikant schlechter als das vollständige Modell, wobei die Varianz insgesamt niedrig ausfiel. Wie aus Tabelle 13.4 hervorgeht lagen die Mittelwerte in allen Motivationsskalen insgesamt in einem hohen Bereich und fielen für K1 in der Regel am höchsten aus. Lediglich für FSK Deutsch sowie IM Mathematik lag zu *Z3* ein etwas höherer Wert in K2 vor. Mit zunehmendem zeitlichen Verlauf fielen die latenten Mittelwerte von FSK und IM niedriger aus. Beim FSK Mathematik sank der Mittelwert in K1 vor allem von *Z2* nach *Z3* etwas stärker ab ($MW_{\text{latent}} = 4.00$ zu *Z1*, $MW_{\text{latent}} = 3.68$ zu *Z3*). Für beide Skalen der IM sanken die latenten Mittelwerte über die Zeit hinweg relativ kontinuierlich in allen Kohorten.

¹ Hier wurde das Modell allerdings nur knapp nicht signifikant, $p < 0.07$.

Tabelle 13.1
Modellprüfung der Veränderung der latenten Mittelwerte und Varianzen – Schulerfolg Mathematik und Deutsch

SE_Ma	Modell	Vergleich	χ^2	df	χ^2/df	CFI	TLI	SRMR	AIC	Δdf	$\Delta\chi^2(\text{TRd})$
K1_AS_2	M1		7.21*	2	3.61	0.99	0.92	0.01	1432.72		
	M2	M1	6.10	4	1.53	0.99	0.98	0.02	1429.40	2	-1.00
	M3	M1	18.09**	4	4.52	0.97	0.89	0.23	1447.85	2	6.59*
K2_AS_2	M1		1.47	2	0.74	1.00	1.01	0.01	2286.63		
	M2	M1	4.76	4	1.19	0.99	0.99	0.04	2285.84	2	3.31
	M3	M1	11.33*	4	2.83	0.98	0.96	0.12	2292.79	2	8.87**
K3_AS_2	M1		10.74**	2	5.37	0.98	0.85	0.01	2144.78		
	M2	M1	12.80**	4	3.20	0.98	0.93	0.03	2141.57	2	2.71
	M3	M1	16.96**	4	4.24	0.97	0.89	0.10	2147.45	2	5.99*
K1_AS_2	M1		1.27	3	0.42	1.00	1.02	0.01	1270.55		
	M2	M1	3.77	4	0.94	1.00	1.00	0.04	1270.37	2	2.70
	M3	M1	3.40	4	0.85	1.00	1.00	0.09	1270.49	2	1.75
K2_AS_2	M1		3.63	3	1.21	1.00	1.00	0.01	1891.95		
	M2	M1	4.61	5	0.92	1.00	1.00	0.04	1889.55	2	0.69
	M3	M1	10.95*	5	2.19	0.99	0.97	0.17	1897.50	2	5.27
K3_AS_2	M1		2.46	3	0.82	1.00	1.01	0.01	1749.21		
	M2	M1	2.82	5	0.56	1.00	1.01	0.02	1745.41	2	0.44
	M3	M1	11.83	5	2.37	0.99	0.96	0.17	1754.12	2	9.81**

Anmerkungen. N(K1_AS_2) = 170; N(K2_AS_2) = 215; N(K3_AS_2) = 190; SE = Schulerfolg; Ma = Mathematik; De = Deutsch; CFI = Comparative Fit Index; TLI = Tucker-Lewis-Index; SRMR = Standardized Root Mean Square Residual; AIC = Akaike Information Criterion; Δdf = Differenz der Freiheitsgrade genesteter Modelle; $\Delta\chi^2(\text{TRd})$ = Satorra-Bentler Chi-Quadrat Differenztest; M1 = Bestfittendes Modell der Messmodellprüfung; M2 = Modell mit gleichgesetzten latenten Mittelwerten; M3 = Modell mit gleichgesetzten latenten Varianzen; Modelle mit signifikanter Verschlechterung im Fit sind *fett* markiert.

* $p < 0.05$. ** $p < 0.01$.

Tabelle 13.2
Modellprüfung der Veränderung der latenten Mittelwerte und Varianzen – FSK Mathematik und Deutsch

	Modell	Vergleich	χ^2	df	χ^2/df	CFI	TLI	SRMR	AIC	Δdf	$\Delta\chi^2(\text{TRd})$
FSK_Ma	K1_AS_2	M1	9.21	15	0.61	1.00	1.04	0.03	3052.77		
	M2	M1	28.25*	17	1.66	0.96	0.92	0.06	3068.59	2	13.34**
	M3	M1	10.38	17	0.61	1.00	1.05	0.04	3049.80	2	1.03
	K2_AS_2	M1	27.39*	15	1.83	0.98	0.94	0.05	3773.37		
		M2	31.86*	17	1.87	0.97	0.94	0.06	3773.90	2	3.92
		M3	41.03**	17	2.41	0.95	0.90	0.14	3782.76	2	15.86**
	K3_AS_2	M1	20.28	15	1.35	0.99	0.97	0.05	3338.82		
		M2	21.73	17	1.28	0.99	0.98	0.05	3336.36	2	1.29
		M3	24.89	17	1.46	0.98	0.96	0.08	3340.17	2	4.04
FSK_De	K1_AS_2	M1	13.29	15	0.89	1.00	1.02	0.04	3015.69		
	M2	M1	16.71	17	0.98	1.00	1.00	0.06	3015.72	2	2.94
	M3	M1	14.01	17	0.82	1.00	1.03	0.06	3012.53	2	0.61
	K2_AS_2	M1	26.76*	15	1.78	0.97	0.94	0.05	4004.42		
		M2	28.34*	17	1.67	0.98	0.95	0.05	4002.13	2	1.06
		M3	27.79*	17	1.63	0.98	0.95	0.05	4003.40	2	0.50
	K3_AS_2	M1	45.59**	15	3.04	0.94	0.86	0.05	3396.11		
		M2	44.48**	17	2.62	0.95	0.89	0.05	3392.49	2	-0.96
		M3	43.13**	17	2.54	0.95	0.89	0.05	3392.47	2	-1.77

Anmerkungen. N(K1_AS_2) = 170; N(K2_AS_2) = 215; N(K3_AS_2) = 190; FSK = Fähigkeitsselbstkonzept; Ma = Mathematik; De = Deutsch; CFI = Comparative Fit Index; TLI = Tucker-Lewis-Index; SRMR = Standardized Root Mean Square Residual; AIC = Akaike Information Criterion; Δdf = Differenz der Freiheitsgrade genesteter Modelle; $\Delta\chi^2(\text{TRd})$ = Satorra-Bentler Chi-Quadrat Differenztest; M1 = Best-fittendes Modell der Messmodellprüfung; M2 = Modell mit gleichgesetzten latenten Mittelwerten; M3 = Modell mit gleichgesetzten latenten Varianzen; Modelle mit signifikanter Verschlechterung im Fit sind *fett* markiert.

* $p < 0.05$.

** $p < 0.01$.

Tabelle 13.3
Modellprüfung der Veränderung der latenten Mittelwerte und Varianzen – IM Mathematik und Deutsch

	Modell	Vergleich	χ^2	df	χ^2/df	CFI	TLI	SRMR	AIC	Δdf	$\Delta\chi^2(\text{TRd})$
IM_Ma	K1_AS_2	M1	7.78	18	0.43	1.00	1.08	0.03	3478.49		
	M2	M1	45.86**	20	2.29	0.90	0.81	0.07	3506.95	2	33.85**
	M3	M1	9.16	20	0.46	1.00	1.08	0.05	3470.15	2	1.04
K2_AS_2	M1		12.69	18	0.71	1.00	1.00	0.03	4402.56		
	M2	M1	29.87	20	1.49	0.98	0.96	0.05	4419.47	2	13.93**
	M3	M1	16.49	20	0.82	1.00	1.02	0.06	4403.73	2	2.45
K3_AS_2	M1		28.92*	18	1.61	0.98	0.95	0.06	3901.59		
	M2	M1	51.76**	20	2.59	0.93	0.88	0.07	3924.27	2	19.84**
	M3	M1	35.34*	20	1.77	0.97	0.94	0.08	3905.51	2	5.01
IM_De	K1_AS_2	M1	23.66**	10	2.37	0.94	0.83	0.06	3118.64		
	M2	M1	25.01*	12	2.08	0.94	0.87	0.06	3115.64	2	1.51
	M3	M1	24.35*	12	2.03	0.95	0.88	0.07	3116.79	2	0.51
K2_AS_2	M1		17.70	10	1.77	0.98	0.94	0.05	4127.37		
	M2	M1	31.74**	12	2.65	0.95	0.88	0.06	4137.21	2	13.71**
	M3	M1	17.75	12	1.48	0.99	0.97	0.05	4123.41	2	0.04
K3_AS_2	M1		8.32	10	0.83	1.00	1.01	0.03	3518.70		
	M2	M1	16.85	12	1.40	0.99	0.97	0.04	3524.99	2	7.53*
	M3	M1	11.69	12	0.97	1.00	1.00	0.06	3518.54	2	3.16

Anmerkungen. N(K1_AS_2) = 170; N(K2_AS_2) = 215; N(K3_AS_2) = 190; IM = Intrinsische Motivation; Ma = Mathematik; De = Deutsch; CFI = Comparative Fit Index; TLI = Tucker-Lewis-Index; SRMR = Standardized Root Mean Square Residual; AIC = Akaike Information Criterion; Δdf = Differenz der Freiheitsgrade genesteter Modelle; $\Delta\chi^2(\text{TRd})$ = Satorra-Bentler Chi-Quadrat Differenztest; M1 = Best-fittendes Modell der Messmodellprüfung; M2 = Modell mit gleichgesetzten latenten Mittelwerten; M3 = Modell mit gleichgesetzten latenten Varianzen; Modelle mit signifikanter Verschlechterung im Fit sind *fett* markiert.

* $p < 0.05$.
 ** $p < 0.01$.

Tabelle 13.4

Latente Mittelwerte und Varianzen von Schulerfolg, Fähigkeitsselbstkonzept und intrinsischer Motivation

		K1_AS_2			K2_AS_2			K3_AS_2		
		Z1	Z2	Z3	Z1	Z2	Z3	Z1	Z2	Z3
SE_Ma	MW	5.00	5.00	5.34	5.20	4.36	4.79	4.61	4.11	4.43
	Varianz	0.37	0.37	0.75	0.53	0.44	0.70	0.43	0.42	0.62
SE_De	MW	4.63	4.73	4.18	4.68	3.77	4.16	4.73	4.68	4.56
	Varianz	0.53	0.45	0.44	0.57	0.33	0.37	0.55	0.40	0.35
FSK_Ma	MW	4.00	4.09	3.68	3.89	3.99	3.83	3.75	3.70	3.64
	Varianz	0.16	0.25	0.24	0.31	0.16	0.48	0.27	0.47	0.51
FSK_De	MW	4.23	4.18	4.32	3.93	4.01	3.97	3.95	3.99	3.97
	Varianz	0.23	0.22	0.16	0.34	0.23	0.21	0.27	0.24	0.29
IM_Ma	MW	4.07	4.21	3.46	3.90	4.05	3.60	3.80	3.77	3.28
	Varianz	0.22	0.15	0.28	0.34	0.20	0.41	0.20	0.22	0.54
IM_De	MW	4.29	4.21	4.23	4.08	3.99	3.74	4.07	4.06	3.82
	Varianz	0.22	0.33	0.19	0.27	0.27	0.24	0.24	0.14	0.31

Anmerkungen. N(K1_AS_2) = 170; N(K2_AS_2) = 215; N(K3_AS_2) = 190; SE = Schulerfolg; FSK = Fähigkeitsselbstkonzept; IM = Intrinsische Motivation; Ma = Mathematik; De = Deutsch; MW = latenter Mittelwert; Latente Mittelwerte und Varianzen, die in der Modellschätzung nicht gleichgesetzt werden konnten sind *fett* markiert.

13.1.2 Interindividuelle Stabilität

Tabelle 13.5 stellt die manifesten sowie latenten Stabilitätskoeffizienten dar. Im Ganzen betrachtet fielen durch die Möglichkeit der Berücksichtigung des Messfehlers die manifesten Stabilitäten niedriger aus als die latenten. Nachfolgend werden für FSK, IM und SE lediglich die latenten Stabilitäten weiter betrachtet, für allgemeine kognitive Fähigkeiten kann aufgrund der fehlgeschlagenen latenten Modellierung zu Z1 nur die manifeste Stabilität inspiziert werden.

Im schulischen Erfolg zeigte sich eine sehr hohe Stabilität, die in Mathematik in K3 von Z2 nach Z3 mit $r_{tt} = 0.88$ am höchsten ausfiel. In beiden Fächern waren die erreichten Stabilitäten insgesamt betrachtet vergleichbar hoch. Bei zeitlich benachbarten Zeitpunkten fanden sich höhere Stabilitäten als zwischen Z1 und Z3. Außerdem zeigte sich, dass die Stabilität in K2 und K3 von Z1 nach Z2 (also über den Wechsel in die Sekundarschule hinweg) niedriger lagen als in K1 im gleichen Zeitintervall. Für K1 lagen die Stabilitäten der Mathematik und Deutschnoten zwischen Z1 und Z2 bei $r_{tt} = 0.74$ bzw. $r_{tt} = 0.79$. Zwischen den Messzeitpunkten Z2 nach Z3 ergab sich eine

Tabelle 13.5

Manifeste und latente Stabilität von Schulerfolg, Fähigkeitsselbstkonzept und intrinsischer Motivation

		Manifest			Latent		
		Z1 - Z2	Z2 - Z3	Z1 - Z3	Z1 - Z2	Z2 - Z3	Z1 - Z3
SE_Ma	K_AS_2	0.48	0.62	0.45	0.63	0.81	0.60
	K1_AS_2	0.45	0.46	0.34	0.74	0.70	0.64
	K2_AS_2	0.50	0.66	0.57	0.62	0.86	0.65
	K3_AS_2	0.41	0.68	0.33	0.51	0.88	0.51
SE_De	K_AS_2	0.49	0.61	0.54	0.63	0.76	0.66
	K1_AS_2	0.78	0.51	0.65	0.79	0.67	0.69
	K2_AS_2	0.47	0.69	0.57	0.64	0.85	0.67
	K3_AS_2	0.44*	0.65	0.45	0.57	0.83	0.65
g-Faktor	K_AS_2			0.41			
	K1_AS_2			0.57			
	K2_AS_2			0.28			
	K3_AS_2			0.46			
FSK_Ma	K_AS_2	0.43	0.57	0.41	0.53	0.61	0.49
	K1_AS_2	0.38	0.47	0.33	0.49	0.49	0.45
	K2_AS_2	0.35	0.58	0.43	0.44	0.61	0.52
	K3_AS_2	0.55	0.67	0.45	0.60	0.71	0.51
FSK_De	K_AS_2	0.51	0.65	0.50	0.57	0.73	0.58
	K1_AS_2	0.47	0.54	0.27	0.61	0.65	0.39
	K2_AS_2	0.46	0.68	0.56	0.49	0.81	0.59
	K3_AS_2	0.54	0.70	0.57	0.56	0.69	0.63
IM_Ma	K_AS_2	0.40	0.51	0.34	0.46	0.58	0.44
	K1_AS_2	0.24	0.47	0.22*	0.24	0.56	0.25
	K2_AS_2	0.44	0.50	0.32	0.52	0.59	0.44
	K3_AS_2	0.48	0.58	0.45	0.46	0.64	0.54
IM_De	K_AS_2	0.48	0.62	0.42	0.51	0.66	0.45
	K1_AS_2	0.40	0.56	0.13 ^{n.s.}	0.54	0.57	0.22*
	K2_AS_2	0.49	0.63	0.47	0.46	0.61	0.47
	K3_AS_2	0.48	0.63	0.52	0.48	0.69	0.54

Anmerkungen. N(K1_AS_2) = 170; N(K2_AS_2) = 215; N(K3_AS_2) = 190; SE = Schulerfolg; FSK = Fähigkeitsselbstkonzept; IM = Intrinsische Motivation; Ma = Mathematik; De = Deutsch; Alle Korrelationen signifikant mit $p < 0.01$, es sei denn explizit anders gekennzeichnet.

* $p < 0.05$. ^{n.s.} nicht signifikant.

höhere Stabilität für Kinder der Kohorten 2 und 3 im Vergleich zu den jüngsten Kindern, die zwischen Z2 und Z3 auf die weiterführende Schule wechselten. Zudem fanden sich auch im längsten Zeitintervall von Z1 nach Z3 beachtliche Stabilitätskoeffizienten in sämtlichen Kohorten. Diese lagen zwischen $r_{tt} = 0.51$ (K3, SE Mathematik) und $r_{tt} = 0.69$ (K2, SE Deutsch).

Für den manifesten g -Faktor konnte gezeigt werden, dass insbesondere in K2 die Stabilität von $Z1$ nach $Z3$ mit $r_{tt} = 0.28$ eher niedrig ausfiel, während sich diese für K1 und K3 in einem moderaten Bereich bewegte ($r_{tt} = 0.57$ bzw. $r_{tt} = 0.46$).

Schließlich konnte für die Stabilitäten der Motivationsmaße FSK und IM insgesamt eine mittlere bis hohe Stabilität konstatiert werden ($r_{tt} = 0.22$ in IM Deutsch, K1, $Z1$ nach $Z3$ bis $r_{tt} = 0.81$ in FSK Deutsch, K2, $Z2$ nach $Z3$). Generell fiel die Stabilität bei aufeinanderfolgenden Messzeitpunkten wiederum höher aus als im längsten Zeitintervall von $Z1$ nach $Z3$. Zudem ließ sich festhalten, dass im FSK höhere Stabilitäten vorlagen als bei der IM. Die höchste Stabilität im FSK lag mit $r_{tt} = 0.81$ in Deutsch für K2 vor ($Z2$ nach $Z3$), in der IM lag die höchste beobachtete Stabilität bei $r_{tt} = 0.69$ (IM Deutsch K3 von $Z2$ nach $Z3$). In den meisten Fällen fiel dabei die Stabilität für die fachspezifischen Skalen in Deutsch etwas höher aus. Außerdem zeigte sich in der Tendenz, dass die Stabilitäten mit steigendem Alter der Kinder höher ausfiel.

13.2 Längsschnittlicher Zusammenhang zwischen den Variablen

13.2.1 Motivation und Schulerfolg

In Bezug auf den Zusammenhang zwischen FSK, IM und SE über den gesamten Untersuchungsverlauf hinweg wurden zunächst die manifesten Skaleninterkorrelationen betrachtet. Dies ermöglichte eine erste Abschätzung der Höhe der Zusammenhänge innerhalb der einzelnen Messzeitpunkte (Querschnitt) sowie zwischen den Messzeitpunkten (Längsschnitt). Potenziell beobachtete Entwicklungstrends auf Basis manifester Skalenergebnisse dienten dabei lediglich als erste Inspektion der Zusammenhänge. Generell sind diese nur unter Vorbehalt zu interpretieren, da hier einerseits einer möglichen Konfundierung mit dem Alter der Kinder keine Rechnung getragen werden kann (Guay et al., 2003) und andererseits keine Berücksichtigung von Messfehlern (für jedes Konstrukt und über die Zeit) möglich ist. Im Vergleich der manifesten und latenten korrelativen Zusammenhänge zeigte sich in der Summe ein übereinstimmendes Muster. Bei latenter Modellierung fielen die Korrelationen jedoch in der Regel etwas höher aus, was auf die Möglichkeit der Berücksichtigung der Messfehler zurückzuführen ist. Nachfolgend werden daher lediglich die latenten Zusammenhänge einer näheren Betrachtung unterzogen. Die entsprechenden manifesten Zusammenhänge können Tabelle B.8 in Anhang B entnommen werden.

Tabelle 13.6 enthält die fachspezifischen latenten Skalenkorrelationen zwischen FSK, IM und SE zu $Z1$, $Z2$ und $Z3$. Für das FSK fielen die Korrelationen in Deutsch in der Regel etwas höher aus als in Mathematik, wobei die *zeitgleichen Korrelationen*

Tabelle 13.6
Latente Skalenerkorrelation zwischen Schulerfolg und Motivation

		Schulerfolg											
		K_AS_2			K1_AS_2			K2_AS_2			K3_AS_2		
		Z1	Z2	Z3	Z1	Z2	Z3	Z1	Z2	Z3	Z1	Z2	Z3
FSK_Ma	Z1	0.60	0.37	0.34	0.50	0.35	0.30	0.62	0.37	0.38	0.67	0.36	0.33
	Z2	0.31	0.39	0.32	0.31	0.41	0.34	0.22	0.28	0.27	0.33	0.44	0.34
	Z3	0.32	0.40	0.43	0.15 ^{n.s.}	0.09 ^{n.s.}	0.24*	0.39	0.42	0.51	0.36	0.55	0.45
FSK_De	Z1	0.68	0.60	0.54	0.59	0.56	0.48	0.72	0.59	0.54	0.78	0.61	0.61
	Z2	0.44	0.61	0.42	0.48	0.65	0.44	0.41	0.56	0.52	0.49	0.55	0.35
	Z3	0.43	0.45	0.44	0.30	0.35	0.42	0.51	0.44	0.48	0.47	0.45	0.42
IM_Ma	Z1	0.40	0.28	0.27	0.27	0.24	0.25	0.39	0.25	0.27	0.50	0.35	0.33
	Z2	0.19	0.30	0.21	0.19*	0.29	0.29	0.18*	0.26	0.20*	0.17 ^{n.s.}	0.30	0.15 ^{n.s.}
	Z3	0.15	0.24	0.30	0.06 ^{n.s.}	0.06 ^{n.s.}	0.26	0.17 ^{n.s.}	0.25*	0.29	0.24*	0.38	0.39
IM_De	Z1	0.42	0.38	0.32	0.36	0.36	0.31	0.37	0.26	0.22*	0.59	0.46	0.44
	Z2	0.27	0.42	0.27	0.25*	0.47	0.32	0.25*	0.32	0.24*	0.39	0.38	0.23 ^{n.s.}
	Z3	0.29	0.34	0.33	0.42 ^{n.s.}	0.55*	0.34^{n.s.}	0.17 ^{n.s.}	0.24 ^{n.s.}	0.33*	0.45	0.36	0.31*

Anmerkungen. N(K1_AS_2) = 170; N(K2_AS_2) = 215; N(K3_AS_2) = 190; FSK = Fähigkeitsselbstkonzept; IM = Intrinsische Motivation; Ma = Mathematik; De = Deutsch; Alle Korrelationen signifikant mit $p < 0.01$, es sei denn explizit anders gekennzeichnet.

* $p < 0.05$. ^{n.s.} nicht signifikant.

(fettgedruckte Werte) bis auf wenige Ausnahmen höher lagen als die *Kreuzkorrelationen* (Korrelation zwischen benachbarten Zeitpunkten). Die zeitgleichen Korrelationen lagen zwischen $r = 0.24$ (K1 zu Z3) und $r = 0.67$ (K3 zu Z1) in Mathematik und zwischen $r = 0.42$ (K1 sowie K3 zu Z3) und $r = 0.78$ (K3 zu Z1) in Deutsch. Insgesamt betrachtet fand sich ein Absinken der Zusammenhänge zwischen FSK und SE über die Zeit. Lediglich in K1 zeigte sich für FSK Deutsch zu Z2 ein höherer Zusammenhang als zu Z1 ($r = 0.65$ vs. $r = 0.59$), dieser fiel zu Z3 jedoch wieder niedriger aus als zu Z1 ($r = 0.42$). In K2 lag zu Z3 ein höherer Zusammenhang zwischen FSK Mathematik und SE als zu Z2 vor ($r = 0.51$ vs. $r = 0.28$). Für das FSK Mathematik konnte festgestellt werden, dass für K2 und K3 von Z1 nach Z2 ein stärkeres Absinken zu verzeichnen war als von Z2 nach Z3. In K1 sanken die Korrelationen von Z2 nach Z3 in stärkerem Ausmaß als zwischen Z1 und Z2. Im FSK Deutsch war dies vergleichbar.

Die Betrachtung der Kreuzkorrelationen lieferte einen ersten Blick auf die Frage nach kausaler Prädominanz zwischen motivationalen Merkmalen und SE.

Für das FSK Mathematik fiel die Beziehung zwischen FSK auf SE in K1 stets höher aus als die umgekehrte Richtung SE auf FSK ($r = 0.35$ und $r = 0.34$ FSK auf SE vs. $r = 0.31$ und $r = 0.09$ SE auf FSK). Wird die Höhe der CL-Beziehung der ersten zur zweiten Erhebung betrachtet, fielen die Korrelationen in beide Richtungen in etwa gleich hoch aus ($r = 0.35$ FSK auf SE vs. $r = 0.31$ SE auf FSK), wohingegen von der zweiten zur dritten Erhebung die Korrelation deutlich zugunsten des FSKs ausfiel. Die Beziehung über das längste Zeitintervall wies eine höhere Korrelation zugunsten des FSKs auf ($r = 0.30$, FSK Z1 nach SE Z3; $r = 0.15$, SE Z1 nach FSK Z3). In K2 zeigten sich in den direkt aufeinanderfolgenden Intervallen Z1 nach Z2 bzw. Z2 nach Z3 moderate CL-Korrelationen zwischen $r = 0.22$ und $r = 0.42$. Im Intervall Z1 nach Z3 lagen vergleichbare Beziehungen zwischen FSK und SE vor ($r = 0.38$ FSK auf SE vs. $r = 0.39$ SE auf FSK). Für K3 zeigte sich zwischen Z1 und Z2 eine vergleichbare Höhe der CL-Korrelation ($r = 0.36$, FSK Z1 nach SE Z2; $r = 0.33$, SE Z1 nach FSK Z2), zwischen Z2 und Z3 lagen die Korrelation von SE auf FSK mit $r = 0.55$ deutlicher höher als die entgegengesetzte Beziehung von FSK auf SE $r = 0.34$. Von Z1 nach Z3 lagen die Korrelationen in beide Richtungen wieder in ungefähr gleicher Höhe.

Im FSK Deutsch fielen die Korrelationen insgesamt etwas höher aus. Hier zeigte sich in K1 im Intervall Z2 nach Z3 eine leicht höhere Korrelation von FSK auf SE ($r = 0.44$, FSK Z2 nach SE Z3; $r = 0.35$, SE Z2 nach FSK Z3). Auch in den übrigen Zeitintervallen dominierte jeweils die Beziehung FSK auf SE. In K2 fand sich diese Prädominanz des FSKs ebenfalls in allen Intervallen. In K3 konnte schließlich eine Prädominanz des SE von Z2 nach Z3 beobachtet werden ($r = 0.45$ vs. $r = 0.35$), während zwischen Z1 und Z2 bzw. Z1 und Z3 der CL-Pfad von FSK auf SE höher ausfiel ($r = 0.61$ vs. $r = 0.49$ bzw. $r = 0.61$ vs. $r = 0.47$).

Wird das Muster der Korrelationen für die Skalen der IM vergleichend betrachtet, fällt zunächst auf, dass die Zusammenhänge niedriger ausfielen und im Gegensatz zum FSK kein eindeutiger Trend in der Höhe der Zusammenhänge im Hinblick auf ein Schulfach vorlag. Mathematik und Deutsch zeigten hier in der Regel ähnlich hohe Zusammenhänge. Die zeitgleichen Korrelationen fielen auch für die IM meist am höchsten aus und lagen zwischen $r = 0.26$ und $r = 0.50$ in Mathematik und zwischen $r = 0.31$ bis $r = 0.59$ in Deutsch. Was die Veränderung der Stärke des Zusammenhangs zwischen IM und SE in beiden Fächern angeht ließ sich analog zum FSK feststellen, dass diese über die Zeit hinweg abnahmen. Eine Ausnahme hiervon fand sich für Deutsch in K1. Hier lag die Korrelation zu Z2 ($r = 0.47$) höher als zu Z1 ($r = 0.36$), fiel nach Z3 ($r = 0.34$) jedoch unter den Wert von Z1.

In der IM Mathematik fielen in K2 und K3 die zeitgleichen Korrelationen zwischen Z1 und Z2 jeweils am stärksten, während diese in K1 konstant blieben. Für Deutsch konnte in K1 und K3 ein stärkeres Absinken des Zusammenhangs nach dem Übergang in die weiterführende Schule beobachtet werden, in K2 war dies in der Tendenz ebenso der Fall, allerdings fielen die Korrelationen hier insgesamt weniger stark.

Auch für die Höhe der Kreuzkorrelationen ließ sich feststellen, dass diese in der Summe etwas niedriger ausfielen als für die Skalen des FSK, wobei die Korrelationen für Deutsch diejenigen von Mathematik wiederum übertrafen. Für Mathematik fanden sich in K1 nahezu identische Korrelationen für beide CL-Beziehungen von Z1 nach Z2, im Intervall von Z2 nach Z3 war die Beziehung von IM auf SE mit $r = 0.29$ höher als die umgekehrte Richtung ($r = 0.06$). In K2 fand sich im Intervall Z1 nach Z2 eine etwas höhere Korrelation für IM auf SE ($r = 0.25$ vs $r = 0.18$). Zwischen Z2 und Z3 drehte sich dies um, hier lag die Korrelation von SE auf IM in einem höheren Bereich ($r = 0.25$ vs. $r = 0.20$). Gleiches war auch für K3 zu beobachten, allerdings war der Unterschied zwischen den Korrelationen hier höher. Für das Intervall von Z1 nach Z3 fand sich schließlich in allen Kohorten eine höhere Korrelation von IM auf SE.

Für IM Deutsch fand sich in K1 zwischen Z1 und Z2 eine höhere CL-Korrelation von vorangegangener IM auf nachfolgenden SE ($r = 0.36$). In den beiden verbleibenden Zeitintervallen war hingegen ein umgekehrter Trend zu beobachten, hier lag die Korrelation von SE auf IM in einem höheren Bereich. In K2 lagen die CL-Beziehungen zwischen den Zeitpunkten in vergleichbarer Höhe (zwischen $r = 0.17$ und $r = 0.26$). In K3 war die Korrelation von vorangegangener IM auf nachfolgenden SE von Z1 nach Z2 höher als die umgekehrte Kreuzkorrelation ($r = 0.46$ vs. $r = 0.39$). Zwischen Z2 und Z3 fiel die Korrelation von SE auf IM hingegen mit $r = 0.36$ höher aus als diejenige von IM auf SE ($r = 0.23$). Im Intervall Z1 nach Z3 lagen vergleichbare CL-Korrelationen vor ($r = 0.44$ bzw. $r = 0.45$).

Insgesamt konnte demnach auf Basis der latenten Korrelationen zunächst kein einheitliches Verursachungsmuster über die Motivationsmaße und Kohorten hinweg festgestellt werden.

13.2.2 Allgemeine kognitive Fähigkeiten, Motivation und Schulerfolg

Neben den manifesten Korrelationen zwischen dem *g*-Faktor, SE und Motivation (vgl. Kapitel 11) wurden auch die latenten Zusammenhänge dieser Variablen betrachtet. Diese sind in Tabelle 13.7 dargestellt.

Tabelle 13.7

Latente Skaleninterkorrelation zwischen g-Faktor zu Z3 und Motivation bzw. Schulerfolg

		<i>g</i> -Faktor (Z3)			
		K_AS_2	K1_AS_2	K2_AS_2	K3_AS_2
SE_Ma	Z1	0.54**	0.62**	0.41**	0.68**
	Z2	0.47**	0.71**	0.39**	0.37**
	Z3	0.45**	0.52**	0.45**	0.41**
SE_De	Z1	0.40**	0.53**	0.17	0.59**
	Z2	0.35**	0.55**	0.24*	0.33**
	Z3	0.45**	0.56**	0.32**	0.50**
FSK_Ma	Z1	0.27**	0.16	0.27	0.40**
	Z2	0.05	0.18	-0.10	-0.01
	Z3	0.20*	0.07	0.12	0.34**
FSK_De	Z1	0.24**	0.22	0.04	0.59**
	Z2	0.12	0.33*	0.01	0.01
	Z3	0.10	-0.07	-0.01	0.39**
IM_Ma	Z1	0.11	0.16	0.04	0.21
	Z2	-0.01	0.20	-0.07	-0.14
	Z3	0.07	-0.05	0.00	0.20*
IM_De	Z1	0.18*	0.20	-0.06	0.42**
	Z2	0.08	0.21	0.04	0.02
	Z3	0.25*	0.08	0.40	0.35

Anmerkungen. N(K1_AS_2) = 170; N(K2_AS_2) = 215;
 N(K3_AS_2) = 190; SE = Schulerfolg; FSK = Fähigkeitsselbstkonzept,
 IM = Intrinsische Motivation; Ma = Mathematik; De = Deutsch.

* $p < 0.05$. ** $p < 0.01$.

Wie hieraus hervorgeht lagen auch in diesem Fall die Korrelationen im Vergleich zur manifesten Korrelation höher (vgl. Tabellen 11.8 und 11.5), blieben aber in Bezug auf das Muster an Zusammenhängen analog zu den berichteten manifesten Zusammenhängen bestehen. Für den Zusammenhang zwischen dem latenten *g*-Faktor (es wurde

nur noch $Z3$ weiter betrachtet) und dem schulischen Erfolg ließen sich durchweg höhere Korrelationen konstatieren, als bei manifester Skalenkorrelation. Diese lagen in einem niedrigen bis hohen Bereich (zwischen $r = 0.17$ und $r = .71$). Die latenten Korrelationen zwischen FSK , IM und dem g -Faktor zu $Z3$ lagen etwas höher als für die manifesten Skalen, blieben aber in insgesamt in einem eher niedrigen bis moderaten Bereich bestehen (höchste Korrelation in $K3$ mit FSK Deutsch $r = 0.59$).

Nach dieser deskriptiven Betrachtung der latenten Zusammenhänge über die Zeit wurden in der Folge längsschnittliche SEM Modelle der Vorhersage von schulischem Erfolg durch Motivation mit und ohne Kontrolle von allgemeinen kognitiven Fähigkeiten geschätzt. Die Ergebnisse dieser Betrachtung innerhalb eines MKSD werden im Folgenden vorgestellt.

13.3 Vorhersage des Schulerfolgs im MKSD

Die Beantwortung der unter 5.2 formulierten längsschnittlichen Fragestellungen erfolgte unter Berücksichtigung des von Marsh et al. (1999) empfohlenen Vorgehens. Es wurden daher zunächst mit der Gesamtstichprobe K_AS_2 die in Abbildung 13.1, 13.3, 13.5 und 13.7 dargestellten Full-Forward Modelle geprüft, die Modellierung der latenten Konstrukte erfolgte dabei auf Basis der in 12.2 bestätigten Messmodelle.

In diesen Full-Forward Modellen werden sowohl die zeitgleichen Korrelationen der Variablen zu den entsprechenden Messzeitpunkten als auch sämtliche Pfade aller Variablen eines Messzeitpunktes auf alle Variablen der nachfolgenden Messzeitpunkte freigeschätzt. Zu $Z2$ und $Z3$ handelte es sich dabei um das Residuum der Variablen² (vgl. 9.3). Zur Beurteilung der kausalen Prädominanz von Motivation und schulischem Erfolg unter Berücksichtigung der Bedeutung von Entwicklungs- und Kohorteneffekten wurden diese Modelle in einem MKSD spezifiziert (vgl. Abbildungen 13.2, 13.4, 13.6 sowie 13.8).

Gegen dieses MKSD Modell wurde in der Folge zunächst die Invarianz der Messmodelle zwischen den Kohorten geprüft. Hiernach erfolgte die Schätzung verschiedener Submodelle mit spezifischen Gleichheitsrestriktionen und deren Vergleich zum vollständigen Modell (= Modell mit der höchsten Invarianzstufe zwischen den Kohorten), um die Bedeutung der CL-Pfade zu beurteilen. Sofern diese Prüfung das Vorliegen einer Veränderung über die Zeit nahelegte, wurde dieser Effekt gemäß T. D. Little et al. (2007) hinsichtlich der Bedeutung von Entwicklungs- und Kohorteneffekten geprüft. Da für dieses Vorgehen mindestens zwei überlappende Messzeitpunkte zwischen den Kohor-

² Die Stabilität und die vorausgegangene Vorhersage sind hierbei kontrolliert.

ten vorausgesetzt werden (vgl. 9.3) wurde K2 aus diesen Analysen ausgeschlossen und lediglich K1 und K3 im überlappenden Zeitfenster (4.Klasse zu 6.Klasse) untersucht. Die Fit-Indikatoren der jeweils getesteten Submodelle sind für das FSK in Tabelle 13.9 und für die IM in Tabelle 13.12 dargestellt. Tabelle 13.10 enthält die Prüfung des Entwicklungseffektes für das FSK Mathematik.

Fähigkeitsselbstkonzept Mathematik

Wie aus Tabelle 13.8 hervorgeht, lieferten die Full-Forward Modelle in der Gesamtstichprobe sowie in den drei Kohorten einen guten bis sehr Modell-Fit, wobei die Anpassung in K1 am besten gelang.

Tabelle 13.8

Full-Forward Modell Fähigkeitsselbstkonzept Mathematik und Deutsch

		χ^2	df	CFI	TLI	SRMR	RMSEA	p	AIC
FSK_Ma	M1	95.41**	62	0.99	0.98	0.04	0.03	0.99	15 973.26
	M1a	72.28	62	0.99	0.98	0.04	0.03	0.86	4449.32
	M1b	90.68**	62	0.98	0.96	0.06	0.05	0.59	5706.51
	M1c	85.98*	62	0.98	0.96	0.05	0.05	0.62	5122.35
FSK_De	M1	83.14*	62	0.99	0.99	0.03	0.02	1.00	15 199.47
	M1a	101.02**	62	0.97	0.94	0.06	0.06	0.19	4230.83
	M1b	79.84*	62	0.99	0.98	0.04	0.04	0.83	5789.84

Anmerkungen. FSK = Fähigkeitsselbstkonzept; Ma = Mathematik; De = Deutsch; CFI = Comparative Fit Index; TLI = Tucker-Lewis-Index; SRMR = Standardized Root Mean Square Residual; RMSEA = Root-Mean-Square-Error of Approximation; p = $P(RMSEA < 0.05)$; AIC = Akaike Information Criterion; M1 = Full Forward Modell für K_AS_2; M1a = Full Forward Modell für K1_AS_2; M1b = Full Forward Modell für K2_AS_2; M1c = Full Forward Modell für K3_AS_2; Keine Modellschätzung für FSK_De in K3 aufgrund des schlechten Fits des Messmodells.

* $p < 0.05$. ** $p < 0.01$.

Abbildung 13.1 zeigt das resultierende vollständige Modell für die Gesamtstichprobe K_AS_2. Die CU, in der Abbildung aus Gründen der Übersichtlichkeit nicht dargestellt, war in den meisten Fällen positiv, jedoch in keinem Fall signifikant negativ. Alle zeitgleichen Korrelationen zwischen FSK und SE waren signifikant und nahmen über die Zeit hinweg ab ($r = 0.61$ zu Z1, $r = 0.24$ zu Z3). Die Autokorrelationen sanken im Fall des FSKs von Z1 nach Z3, während diese für SE leicht anstiegen. Die kritischen Pfade zur Beurteilung der kausalen Prädominanz von FSK und SE stellen diejenigen Pfade dar, die von den Z1 Konstrukten zu den Z2 Konstrukten bzw. von den Z2 Konstrukten zu den Z3 Konstrukten führen. Der einzige signifikante Pfad stellte hier derjenige von Z2 SE nach Z3 FSK dar ($\beta = 0.17$), beide Beziehungsrichtungen von Z1 nach Z2 wiesen auf keinerlei Beeinflussung hin ($\beta = 0.00$). Dieses Muster

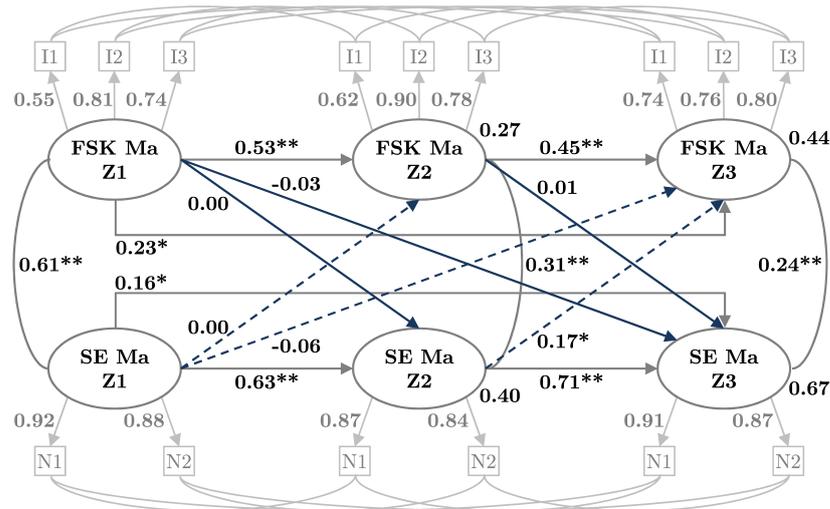


Abbildung 13.1. Full-Forward Modell Fähigkeitsselfkonzept Mathematik.
 $*p < 0.05$; $**p < 0.01$.

wies dementsprechend eher auf die Bestätigung eines *skill-development* Modells hin. Unter Berücksichtigung der unter Kapitel 3.4.1 dargestellten Überlegungen in Bezug auf mögliche Entwicklungstrends in der Beeinflussung von SE durch FSK konnte vermutet werden, dass sich innerhalb der Kohorten ein hiervon differierendes Muster der Beziehungen finden ließe. Aus diesem Grund wurde in der Folge das Full-Forward Modell auch in einem MKSD umgesetzt. Die Ergebnisse dieser Modellierung werden im Folgenden im Detail vorgestellt.

Wie aus Abbildung 13.2³ hervorgeht, zeigten sich bei Betrachtung der einzelnen Kohorten durchaus unterschiedliche Beziehungsmuster als in der Gesamtstichprobe, wobei die meisten Pfade aufgrund der geringen Stichprobengröße und der Korrektur um die Abhängigkeit der Stichprobe nicht signifikant wurden. Die Höhe der Pfadschätzungen kann dennoch interpretiert werden.

In K1 überwog zwischen Z1 und Z2 der Einfluss von vorangegangenen SE auf nachfolgendes FSK ($\beta = 0.11$), wobei dieser Pfad in einem niedrigen Bereich lag. Zwischen Z2 und Z3, also von der 4. zur 6. Klasse, fand sich ein negativer Pfad in Höhe von $\beta = -0.20$ von SE auf FSK, der umgekehrte CL-Pfad lag bei $\beta = 0.10$. Bessere Noten in Mathematik in der 4. Klasse gingen dementsprechend mit einem niedrigeren FSK in der 6. Klasse einher. In K2 und K3 fand sich im Zeitintervall Z1 nach Z2 – welches den Übergang in die weiterführende Schule umfasste – ebenfalls ein negatives Pfadgewicht von SE auf FSK ($\beta = -0.11$ in K2 und $\beta = -0.21$ in K3). Für K3 fand sich zudem eine Beeinflussung von FSK auf SE für dieses Intervall ($\beta = 0.16$). Im Zeitintervall Z2 nach

³ Dargestellt werden in diesen Modellen nur noch die interessierenden CL und Stabilitätspfade, auf die Darstellung der zeitgleichen Korrelation sowie der CU wurde aus Gründen einer besseren Lesbarkeit der Abbildung verzichtet.

Z3 lag in beiden Kohorten ein positives Pfadgewicht von SE auf FSK vor ($\beta = 0.16$ in K2 und $\beta = 0.33^{**}$ in K3), der umgekehrte CL-Pfad von FSK nach SE war hingegen unbedeutend ($\beta = 0.02$ in K2 und $\beta = -0.03$ in K3).

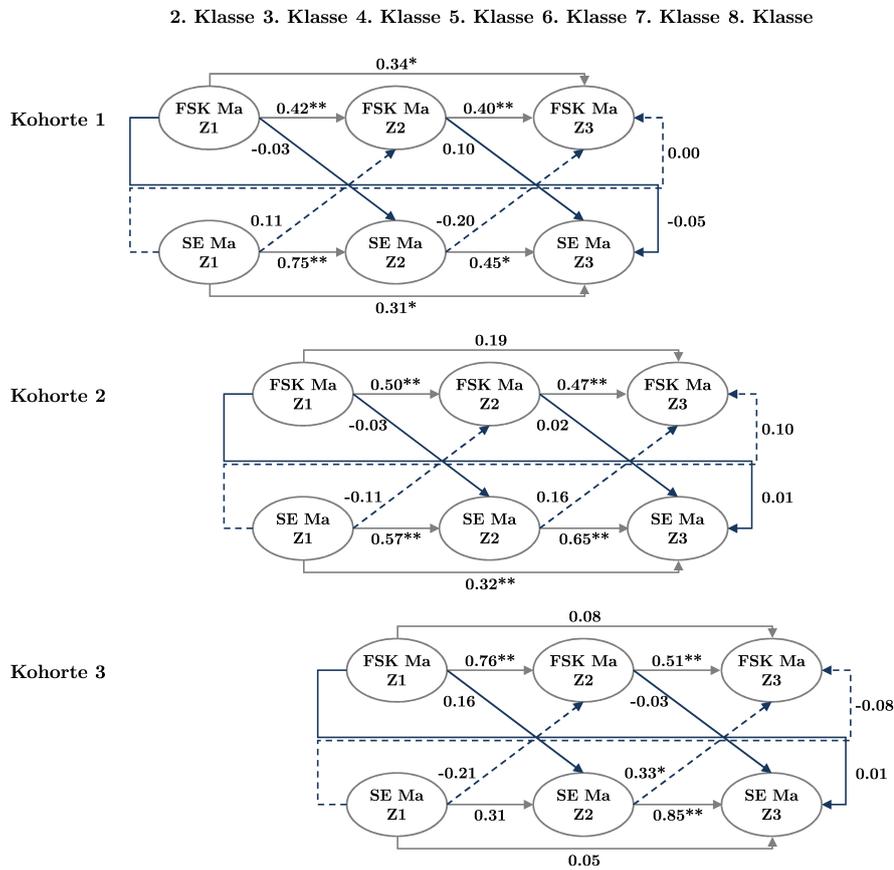


Abbildung 13.2. Multi-Kohorten Modell Fähigkeitsselbstkonzept Mathematik. * $p < 0.05$; ** $p < 0.01$.

Diese für jüngere und ältere Kinder differierenden Beeinflussungsmuster in den Zeitintervallen Z1–Z2 bzw. Z2–Z3 sprachen für das Vorliegen von Entwicklungseffekten, was in der Folge durch Modellvergleiche im MKSD geprüft wurde. Die resultierenden Modellschätzungen können Tabelle 13.9 entnommen werden.

Es wurde zunächst die Invarianz der Faktorladungen, Faktorvarianzen, Intercepts und korrelierten Fehler auf Invarianz zwischen den Kohorten geprüft, hiernach wurde zusätzlich die Stabilität im FSK sowie des SEs zwischen den Kohorten gleichgesetzt (Marsh et al., 1998). In *Modell 2* wurden sämtliche Parameter zwischen den Kohorten freigesetzt. Dieses Modell zeigte eine sehr gute Anpassung an die empirischen Daten. In *Modell 3* wurden die Faktorladungen in *Modell 4* zusätzlich die Faktorvarianzen zwischen den Kohorten gleichgesetzt. Beide Modelle zeigten keine signifikante Verschlechterung des Modell-Fits im Vergleich zu *Modell 2* bzw. *Modell 3*. *Modell 5* enthielt

Tabelle 13.9

Modellfit im MKSD – Fähigkeitsselbstkonzept Mathematik und Deutsch

	Modell	Vergleich	χ^2	df	CFI	TLI	SRMR	RMSEA	p	AIC	Δ df	$\Delta\chi^2(TRd)$
FSK_Ma	M2		242.24**	186	0.98	0.97	0.05	0.04	0.89	15835.35		
	M3	M2	261.34**	204	0.98	0.98	0.06	0.04	0.93	15824.85	18	14.86
	M4	M3	285.37**	216	0.98	0.97	0.11	0.04	0.88	15830.87	12	17.75
	M5	M4	374.69**	246	0.96	0.95	0.15	0.05	0.35	15870.05	30	77.32**
	M6	M5	402.89**	272	0.96	0.96	0.15	0.05	0.48	15858.28	26	21.25
	M6a	M4	314.60**	242	0.98	0.97	0.11	0.04	0.92	15818.65	26	22.20
	M7	M6a	320.44**	248	0.98	0.97	0.12	0.04	0.93	15812.68	6	5.56
	M8	M7	321.93**	254	0.98	0.98	0.11	0.04	0.96	15806.81	6	0.88
	M9	M8	343.13**	263	0.98	0.97	0.13	0.04	0.93	15812.06	9	19.19*
	M10	M8	324.89**	263	0.98	0.98	0.10	0.04	0.98	15794.06	9	2.31
FSK_De	M2		182.30**	124	0.98	0.96	0.05	0.05	0.50	10020.67		
	M3	M2	183.89**	133	0.98	0.97	0.05	0.04	0.75	10006.50	9	1.11
	M4	M3	194.79**	139	0.98	0.97	0.07	0.05	0.67	10010.88	6	8.61
	M5	M4	271.17**	154	0.95	0.94	0.11	0.06	0.04	10065.63	15	61.59**
	M6	M5	290.56**	167	0.95	0.94	0.13	0.06	0.05	10064.61	13	15.43
	M6a	M4	214.55**	152	0.98	0.97	0.08	0.05	0.65	10009.41	13	15.59
	M7	M6a	215.46**	155	0.98	0.97	0.09	0.05	0.70	10006.37	3	0.60
	M8	M7	229.23**	158	0.97	0.96	0.09	0.05	0.56	10013.93	3	18.19**
	M9a	M7	228.57**	161	0.97	0.97	0.10	0.05	0.64	10009.56	6	10.86
	M10a	M7	224.45**	161	0.98	0.97	0.10	0.05	0.70	10004.01	6	8.38

Anmerkungen. CFI = Comparative Fit Index; TLI = Tucker-Lewis-Index; SRMR = Standardized Root Mean Square Residual; RMSEA = Root-Mean-Square-Error of Approximation; p = $P(RMSEA < 0.05)$; AIC = Akaike Information Criterion; Δ df = Differenz der Freiheitsgrade genesteter Modelle; $\Delta\chi^2(TRd)$ = Satorra-Bentler Chi-Quadrat Differenztest; M2 = M1 der Full-Forward Modelle, alle Parameter zwischen den Kohorten frei; M3 = Faktorladungen zwischen den Kohorten gleichgesetzt; M4 = M3 und Faktorvarianzen zwischen den Kohorten; M5 = M4 und Intercepts zwischen den Kohorten; M6 = M5 und korrelierte Fehler zwischen den Kohorten; M6a = M4 und korrelierte Fehler zwischen den Kohorten; M7 = M6a und Stabilität FSK zwischen den Kohorten gleichgesetzt; M8 = M7 und Stabilität SE zwischen den Kohorten gleichgesetzt. M9 = M8 und Pfade von SE auf FSK auf 0 fixiert; M10 = M8 und Pfade von FSK auf SE auf 0 fixiert; M9a = M7 und Pfade von SE auf FSK auf 0 fixiert; M10a = M7 und Pfade von FSK auf SE auf 0 fixiert. * $p < 0.05$. ** $p < 0.01$.

Tabelle 13.10
 Modelle zur Beurteilung eines Entwicklungstrends – FSK Mathematik

	Modell	Vergleich	χ^2	df	CFI	TLI	SRMR	RMSEA	p	AIC	Δ df	$\Delta\chi^2(TRd)$
FSK_Ma	M1		201.73 ^{**}	158	0.98	0.97	0.11	0.04	0.87	9838.86		
	M2	M1	204.79 [*]	164	0.98	0.98	0.09	0.04	0.91	9831.96	6	2.30
	M3	M2	215.20^{**}	167	0.98	0.97	0.10	0.04	0.86	9837.79	3	8.85[*]
	M4	M2	205.66 [*]	165	0.98	0.98	0.10	0.04	0.91	9830.95	1	0.78
											3	8.08 [*]

Anmerkungen. CFI = Comparative Fit Index; TLI = Tucker-Lewis-Index; SRMR = Standardized Root Mean Square Residual; RMSEA = Root-Mean-Square-Error of Approximation; p = $P(RMSEA < 0.05)$; AIC = Akaike Information Criterion; Δ df = Differenz der Freiheitsgrade genesteter Modelle; $\Delta\chi^2(TRd)$ = Satorra-Bentler Chi-Quadrat Differenztest; M1 = M7 des MKSD Modells mit K1 und K3; M2 = M1 und Pfade von FSK auf SE auf 0 fixiert; M3 = M2 und Pfade von SE auf FSK innerhalb und zwischen den Kohorten gleichgesetzt; M4 = Abschätzung des Kohorteneffektes, Pfade im überlappenden Zeitfenster gleichgesetzt; Abnahme des Fits der aus dieser Restriktion resultiert, bildet den Kohorteneffekt ab. Verbleibende Abnahme im Fit von M2 zu M3 kann als Entwicklungseffekt interpretiert werden.

* $p < 0.05$.

** $p < 0.01$.

zusätzlich gleichgesetzte Intercepts zwischen den Kohorten. Dieses Modell zeigte einen signifikant schlechteren Modell-Fit als *Modell 4*, die Intercepts konnten folglich nicht als invariant zwischen den Kohorten angenommen werden. In *Modell 6* wurden schließlich die korrelierten Fehler zwischen den Gruppen gleichgesetzt. Hieraus resultierte kein signifikant schlechterer Fit als für *Modell 5*. Letztlich wurden die Intercepts wieder freigesetzt und das resultierende Modell (*Modell 6a*) als sparsamstes Modell angenommen. In einem nächsten Schritt wurde das Vorherrschen eines *skill-development*, *self-enhancement* oder *REM* Modells in unterschiedlichen Altersbereichen geprüft. Es wurde zunächst in *Modell 7* und *Modell 8* die Stabilität von FSK und SE über die Kohorten hinweg gleichgesetzt (vgl. Marsh et al., 1998). Beide Modelle zeigten einen guten Modell-Fit und im Vergleich zu *Modell 6a* keine signifikante Verschlechterung des Fits. Die Stabilitätskoeffizienten konnten also in allen drei Kohorten als vergleichbar angenommen werden. In *Modell 9* und *Modell 10* wurden nun die interessierenden CL-Pfade von SE auf FSK bzw. von FSK auf SE auf Null fixiert und auf Unterschiedlichkeit zu *Modell 8* geprüft. Wie aus Tabelle 13.9 hervorgeht, resultierte eine Fixierung der Pfade von SE auf FSK in einer signifikanten Abnahme des Modell-Fits. Es lag dementsprechend nahe, hier von einem Entwicklungstrend auszugehen.

Mit dem unter Kapitel 9.3 dargestellten Verfahren von T. D. Little et al. (2007) kann dieser innerhalb eines MKSD von einem Kohorteneffekt getrennt und in seiner Höhe abgeschätzt werden. Wie zuvor bereits dargelegt, wurden hierfür K1 und K3 herangezogen, die Ergebnisse dieser Modellprüfung sind Tabelle 13.10 zu entnehmen. Hier wurden zunächst die *Modelle 8, 9 und 10* der vorangegangenen Betrachtung in einem Multigruppen Modell mit K1 und K3 geprüft. Das Ergebnis dieser Modelle (*Modell 1, 2 und 3*) war das Gleiche wie zuvor, der Fit war auch hier als gut zu bezeichnen. In *Modell 4* wurden zur Abschätzung des Kohorten- und Entwicklungseffektes der interessierenden Beziehung im überlappenden Zeitfenster (SE auf FSK, 4. zur 6. Klasse) zwischen den Kohorten gleichgesetzt. Die resultierende Verringerung im Modell-Fit im Vergleich zu *Modell 2* kann als Kohorteneffekt interpretiert werden, die restliche Abnahme des Fits von *Modell 3* im Vergleich zu *Modell 2* bildet den Entwicklungseffekt ab. Der Kohorteneffekt wurde dementsprechend nicht signifikant.

Fähigkeitsselbstkonzept Deutsch

Für das FSK Deutsch wurden die Modelle aufgrund der geringen Güte des Messmodells für K3 (vgl. Kapitel 12.2) lediglich unter Berücksichtigung von K1 und K2 durchgeführt, eine Abschätzung von potenziellen Entwicklungs- und Kohorteneffekten konnte dementsprechend nicht durchgeführt werden. Tabelle 13.8 enthält die Modellschätzung des Full-Forward Modells (vgl. Abbildung 13.3) für K1 und K2. Der Fit lag insgesamt in einem moderaten Bereich und fiel für K2 besser aus als für K1. Die resultieren-

den Parameterschätzungen innerhalb der MKSD Modellierung sind in Abbildung 13.4 dargestellt.

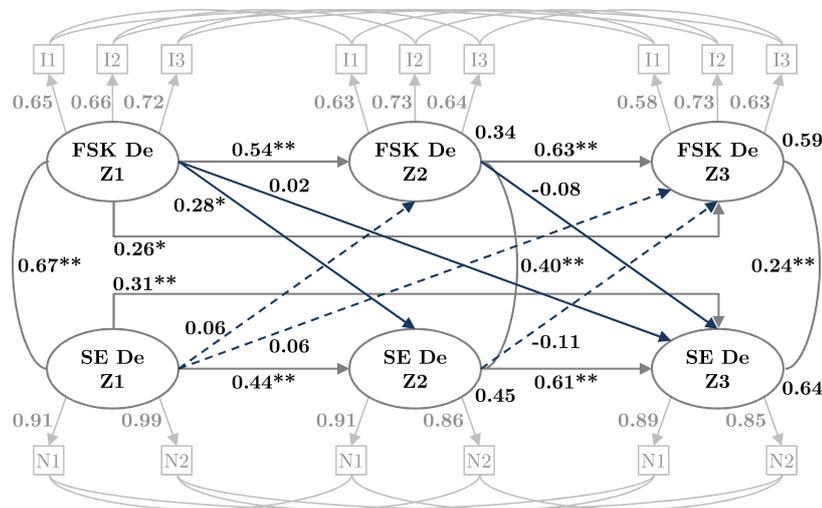


Abbildung 13.3. Full-Forward Modell Fähigkeitsselbstkonzept Deutsch.
* $p < 0.05$; ** $p < 0.01$.

In beiden Kohorten sanken die zeitgleichen Korrelationen über den Untersuchungszeitraum hinweg. Die Pfadschätzungen der CL-Pfade erreichten auch hier meist keine Signifikanz. In K1 zeigte sich von Z1 nach Z2 vor allem ein Effekt von SE auf FSK ($\beta = 0.19$). Im Intervall Z2 nach Z3 fand sich ein negatives Pfadgewicht von SE auf FSK in Höhe von $\beta = -0.13$. In K2 lag ein mittlerer Effekt des FSKs zu Z1 auf SE zu Z2 vor ($\beta = 0.26$), wohingegen zwischen Z2 und Z3 ein negatives Pfadgewicht von SE auf FSK festgestellt werden konnte ($\beta = -0.23$).

Die Modellprüfung erfolgte analog zum FSK Mathematik und kann ebenfalls Tabelle 13.9 entnommen werden. Wie hieraus hervorgeht war auch in diesem Fall *Modell 6a* dasjenige Invarianzmodell mit einem nicht signifikant schlechteren Fit als das vollständige Modell. Im Gegensatz zum FSK in Mathematik konnte für FSK Deutsch allerdings nicht von einer vergleichbaren Stabilität schulischen Erfolgs in den Kohorten ausgegangen werden, da *Modell 8* signifikant schlechter fittete. Die Modelle zur Überprüfung eines Entwicklungstrends (*Modelle 9a* und *10a*) wurden entsprechend mit *Modell 7* verglichen. Tendenziell zeigte sich hier ein zum FSK Mathematik korrespondierendes Muster, denn *Modell 9a* fittete auf einem 10% Niveau signifikant schlechter ($p = 0.09$). Eine weitere Prüfung konnte an dieser Stelle nicht vorgenommen werden.

Intrinsische Motivation Mathematik

Für die IM Mathematik zeigten die Full-Forward Modelle ebenfalls eine gute bis sehr gute Anpassung an die Daten, dabei lag wiederum der beste Fit in K1 vor (vgl. Tabelle

2. Klasse 3. Klasse 4. Klasse 5. Klasse 6. Klasse 7. Klasse 8. Klasse

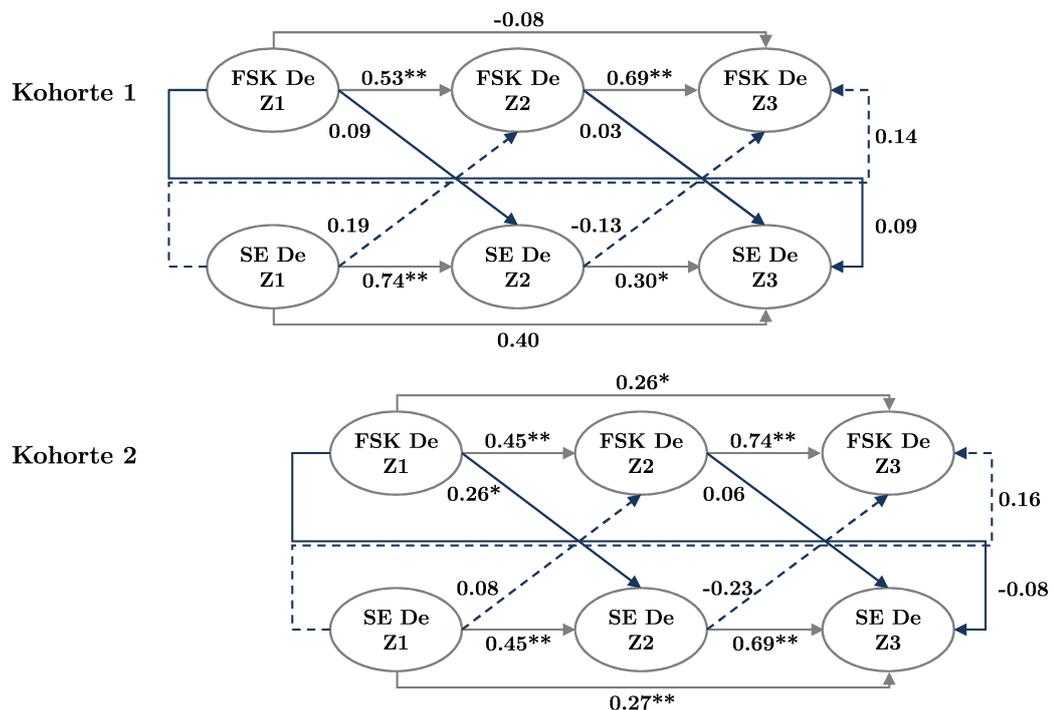


Abbildung 13.4. Multi-Kohorten Modell Fähigkeitsselfkonzept Deutsch.
 $*p < 0.05$; $**p < 0.01$.

13.11). Das resultierende Modell der Gesamtstichprobe K_AS_2 wird in Abbildung 13.5 dargestellt. Wie hieraus hervorgeht lagen die CL-Pfade in beide Verursachungsrichtungen in einem niedrigen Bereich ($\beta = 0.08$ als höchste Pfadschätzung). Die zeitgleichen Korrelationen sanken von Z1 nach Z2 und blieben dann nach Z3 in vergleichbarer Höhe bestehen.

Ähnlich wie bereits im FSK Mathematik beobachtet, lieferte die Betrachtung der Zusammenhänge in den Kohorten einen tieferen Einblick in die Beziehungsmuster. Wie Abbildung 13.6 entnommen werden kann, waren vor allem in K2 niedrige CL-Pfade zu beobachten. In K1 lag im Zeitintervall Z1 nach Z2 ein Effekt von SE auf IM vor ($\beta = 0.12$) im darauf folgenden Intervall fand sich ein negativer Effekt von SE auf IM ($\beta = -0.17$) und ein positives Pfadgewicht von $\beta = 0.12$ von IM auf SE. In K3 fand sich ebenfalls von der 4. zur 6. Klasse ein positiver Effekt von IM auf SE ($\beta = 0.12$) und ein negativer, wenngleich weniger starker, Effekt von SE auf IM ($\beta = -0.09$). Zwischen Z2 und Z3 lag ein negativer Effekt von IM auf SE und ein positiver von SE auf IM vor ($\beta = -0.12$ bzw. $\beta = 0.18$).

Die in Tabelle 13.12 zusammengefassten Modellschätzungen im MKSD ergaben, dass die Intercepts und Varianzen zwischen den Kohorten nicht gleichgesetzt werden konnten. Modell 6b stellt damit das sparsamste Modell dar. Gegen dieses wurde die Folge-

Tabelle 13.11

Full-Forward Modell intrinsische Motivation Mathematik und Deutsch

		χ^2	df	CFI	TLI	SRMR	RMSEA	p	AIC
IM_Ma	M1	86.55*	62	0.99	0.99	0.04	0.03	1.00	17 683.62
	M1a	72.28	62	0.99	0.98	0.04	0.03	0.86	4449.32
	M1b	73.18	62	0.99	0.99	0.05	0.03	0.92	6674.24
	M1c	96.84**	62	0.97	0.95	0.07	0.06	0.34	6001.05
IM_De	M1	91.56**	51	0.99	0.98	0.05	0.04	0.96	15 752.62
	M1a	104.87**	51	0.95	0.91	0.07	0.08	0.02	4371.19
	M1b	69.13*	51	0.99	0.97	0.06	0.04	0.72	6006.49
	M1c	63.28	51	0.99	0.98	0.05	0.04	0.79	5228.93

Anmerkungen. IM = Intrinsische Motivation; Ma = Mathematik; De = Deutsch; CFI = Comparative Fit Index; TLI = Tucker-Lewis-Index; SRMR = Standardized Root Mean Square Residual; RMSEA = Root-Mean-Square-Error of Approximation; p = $P(RMSEA < 0.05)$; AIC = Akaike Information Criterion; M1 = Full Forward Modell für K_AS_2; M1a = Full Forward Modell für K1_AS_2; M1b = Full Forward Modell für K2_AS_2; M1c = Full Forward Modell für K3_AS_2.

* $p < 0.05$. ** $p < 0.01$.

modelle zur Prüfung eines Entwicklungstrends getestet. Es zeigte sich weiterhin, dass keines dieser Modelle (*Modell 7, 8, 9 und 10*) zu einer signifikanten Verringerung des Modell-Fits führte, weshalb an dieser Stelle keine weitere Analyse zur Identifikation von Entwicklungs- und Kohorteneffekten durchgeführt wurde.

Intrinsische Motivation Deutsch

Als letztes Vorhersagemodell wurde schließlich die IM Deutsch betrachtet. Wie Abbildung 13.7 für das Full-Forward Modell der Gesamtstichprobe zeigt, sanken auch in der IM Deutsch die zeitgleichen Korrelationen über die Zeit hinweg. Die CL-Pfade lagen insgesamt, ebenso wie bereits in der IM Mathematik beobachtet, in einem niedrigen Bereich. Lediglich der CL-Pfad von IM zu Z1 nach SE zu Z2 lag mit $\beta = 0.13^*$ etwas höher.

Der Fit der Full-Forward Modelle lag in der Gesamtstichprobe sowie den Kohorten in einem guten bis sehr guten Bereich, hier zeigte sich die schlechteste Passung in K1 (vgl. Tabelle 13.11).

Die Betrachtung der Full-Forward Modelle in den einzelnen Kohorten (vgl. Abbildung 13.8) zeigte in K1 nur geringe Pfadschätzungen in beide Verursachungsrichtungen. Hier lag die Pfadschätzung von SE zu Z1 auf IM zu Z3 mit $\beta = 0.18$ am höchsten. Auch in K2 lagen die Pfadgewichte nahe Null (Ausnahme: SE Z2 nach IM Z3 mit $\beta = -0.10$). In K3 waren die deutlichsten Zusammenhänge von IM auf SE zu finden. Diese lag von Z1 nach Z2 bei $\beta = 0.17$ und von Z2 nach Z3 bei $\beta = -0.18$. Die Einflussrichtung SE

Tabelle 13.12
Modellfit im MKSD – Intrinsicische Motivation Mathematik und Deutsch

	Modell	Vergleich	χ^2	df	CFI	TLI	SRMR	RMSEA	p	AIC	Δ df	$\Delta\chi^2(TRd)$
IM_Ma	M2		242.45**	186	0.98	0.97	0.06	0.04	0.89	17570.51		
	M3	M2	268.86**	204	0.98	0.97	0.06	0.04	0.88	17567.55	18	20.97
	M4	M3	297.84**	216	0.98	0.96	0.11	0.05	0.76	17578.10	12	21.66*
	M5	M4	390.62**	246	0.96	0.95	0.15	0.06	0.19	17620.29	30	80.29**
	M6	M5	414.05**	272	0.96	0.95	0.15	0.05	0.34	17598.50	26	19.56
	M6b	M4	295.28**	230	0.98	0.97	0.07	0.04	0.94	17547.47	26	22.18
	M7	M6b	303.16**	236	0.98	0.97	0.07	0.04	0.94	17546.66	6	5.61
	M8	M7	308.23**	242	0.98	0.97	0.08	0.04	0.95	17543.49	6	3.29
	M9	M8	313.68**	251	0.98	0.98	0.08	0.04	0.97	17531.74	9	4.96
	M10	M8	311.69**	251	0.98	0.98	0.08	0.04	0.98	17530.02	9	3.07
IM_De	M2		235.56**	153	0.98	0.96	0.06	0.05	0.34	15606.61		
	M3	M2	237.38**	169	0.98	0.97	0.07	0.05	0.67	15593.84	16	1.04
	M4	M3	257.02**	181	0.98	0.97	0.08	0.05	0.64	15593.38	12	16.70
	M5	M4	390.08**	209	0.95	0.93	0.13	0.07	0.00	15686.09	28	109.01**
	M6	M5	424.57**	231	0.94	0.93	0.14	0.07	0.00	15681.22	22	30.11
	M6a	M4	289.74**	203	0.97	0.97	0.09	0.05	0.63	15586.58	22	28.21
	M7	M6a	297.06**	209	0.97	0.97	0.09	0.05	0.65	15583.01	6	6.28
	M8	M7	314.32**	215	0.97	0.96	0.09	0.05	0.53	15591.07	6	13.88*
	M9a	M7	302.90**	218	0.98	0.97	0.10	0.05	0.74	15572.85	9	4.93
	M10a	M7	303.88**	218	0.98	0.97	0.10	0.05	0.73	15573.50	9	5.87

Anmerkungen. CFI = Comparative Fit Index; TLI = Tucker-Lewis-Index; SRMR = Standardized Root Mean Square Residual; RMSEA = Root-Mean-Square-Error of Approximation; p = $P(RMSEA < 0.05)$; AIC = Akaike Information Criterion; Δ df = Differenz der Freiheitsgrade genesteter Modelle; $\Delta\chi^2(TRd)$ = Satorra-Bentler Chi-Quadrat Differenztest; M2 = M1 der Full-Forward Modelle, alle Parameter zwischen den Kohorten frei; M3 = Faktorladungen zwischen den Kohorten gleichgesetzt; M4 = M3 und Faktorvarianzen zwischen den Kohorten; M5 = M4 und Intercepts zwischen den Kohorten; M6 = M5 und korrelierte Fehler zwischen den Kohorten; M6a = M4 und korrelierte Fehler zwischen den Kohorten; M6b = M3 und korrelierte Fehler zwischen den Kohorten; M7 = M6a und Stabilität FSK zwischen den Kohorten gleichgesetzt; M8 = M7 und Stabilität SE zwischen den Kohorten gleichgesetzt. M9 = M8 und Pfade von SE auf FSK auf 0 fixiert; M10 = M8 und Pfade von FSK auf SE auf 0 fixiert; M9a = M7 und Pfade von SE auf FSK auf 0 fixiert; M10a = M7 und Pfade von FSK auf SE auf 0 fixiert.
* $p < 0.05$. ** $p < 0.01$.

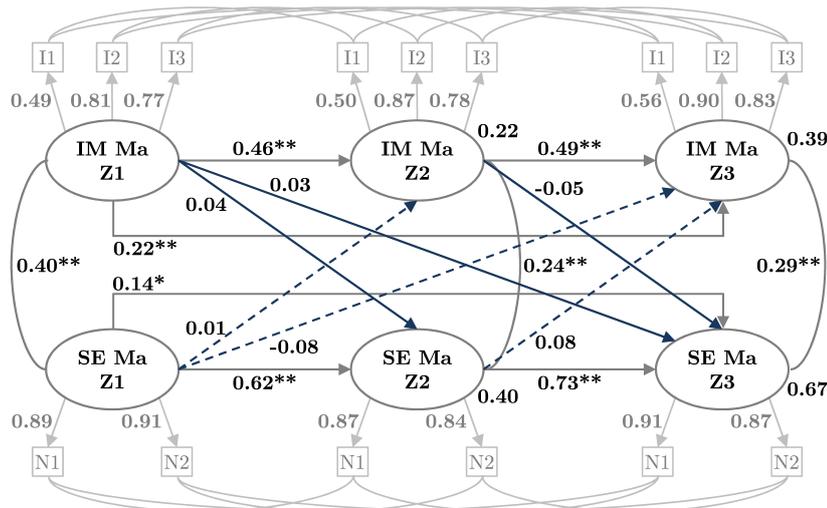


Abbildung 13.5. Full-Forward Modell Intrinsische Motivation Mathematik.
 $*p < 0.05$; $**p < 0.01$.

auf IM war hingegen nur wenig bedeutsam. Dieses Bild bestätigte die durchgeführte Modellschätzung im MKSD. Hier konnten wie bereits beim FSK Deutsch die Intercepts (*Modell 5*) sowie die Stabilität des SEs (*Modell 8*) nicht als invariant angenommen werden. Die Fixierung der CL-Pfade auf Null (*Modelle 9a* und *10a*) führte jeweils zu keiner signifikanten Veränderung im Modell-Fit. Auf eine Analyse von Entwicklungs- und Kohorteneffekten wurde demnach verzichtet.

Wie aus der Darstellung der Modell-Fit Analysen hervorgeht, zeigte sich lediglich im FSK Mathematik ein Entwicklungseffekt über den Untersuchungszeitraum. Für diese motivationale Variable wurde in der Folge geprüft, ob die gefundenen Beziehungen zwischen SE und FSK auch bei Kontrolle der allgemeinen kognitiven Fähigkeiten bestehen bliebe. Es wurde daher in einem weiteren Schritt der latente g -Faktor des dritten Messzeitpunktes in das MKSD Modell aufgenommen. Dabei wurde eine Korrelation zwischen g und SE zu jedem Zeitpunkt zugelassen, eine Korrelation zum FSK wurde hingegen auf Grundlage der latenten Skaleninterkorrelationen nicht spezifiziert.

Gemeinsame Betrachtung von FSK Mathematik und g -Faktor

Im Vergleich der Modelle mit und ohne Kontrolle von g auf Kohortenebene (vgl. Abbildung 13.9; der Fit des Full-Forward Modells findet sich in Tabelle B.9 in Anhang B) zeigte sich für K2 keine wesentliche Veränderung in den Pfadschätzungen. In K1 und K3 blieb das Muster der Zusammenhänge ebenfalls bestehen, allerdings sanken die Autokorrelationen von SE zwischen Z1 und Z2 vor allem in K1 beträchtlich. Dies hatte zur Folge, dass sich die Höhe CL-Pfade im Intervall Z1 nach Z2 ebenfalls veränderte. Der Einfluss sowohl von SE auf FSK als auch von FSK auf SE wurde in K3 schwächer

2. Klasse 3. Klasse 4. Klasse 5. Klasse 6. Klasse 7. Klasse 8. Klasse

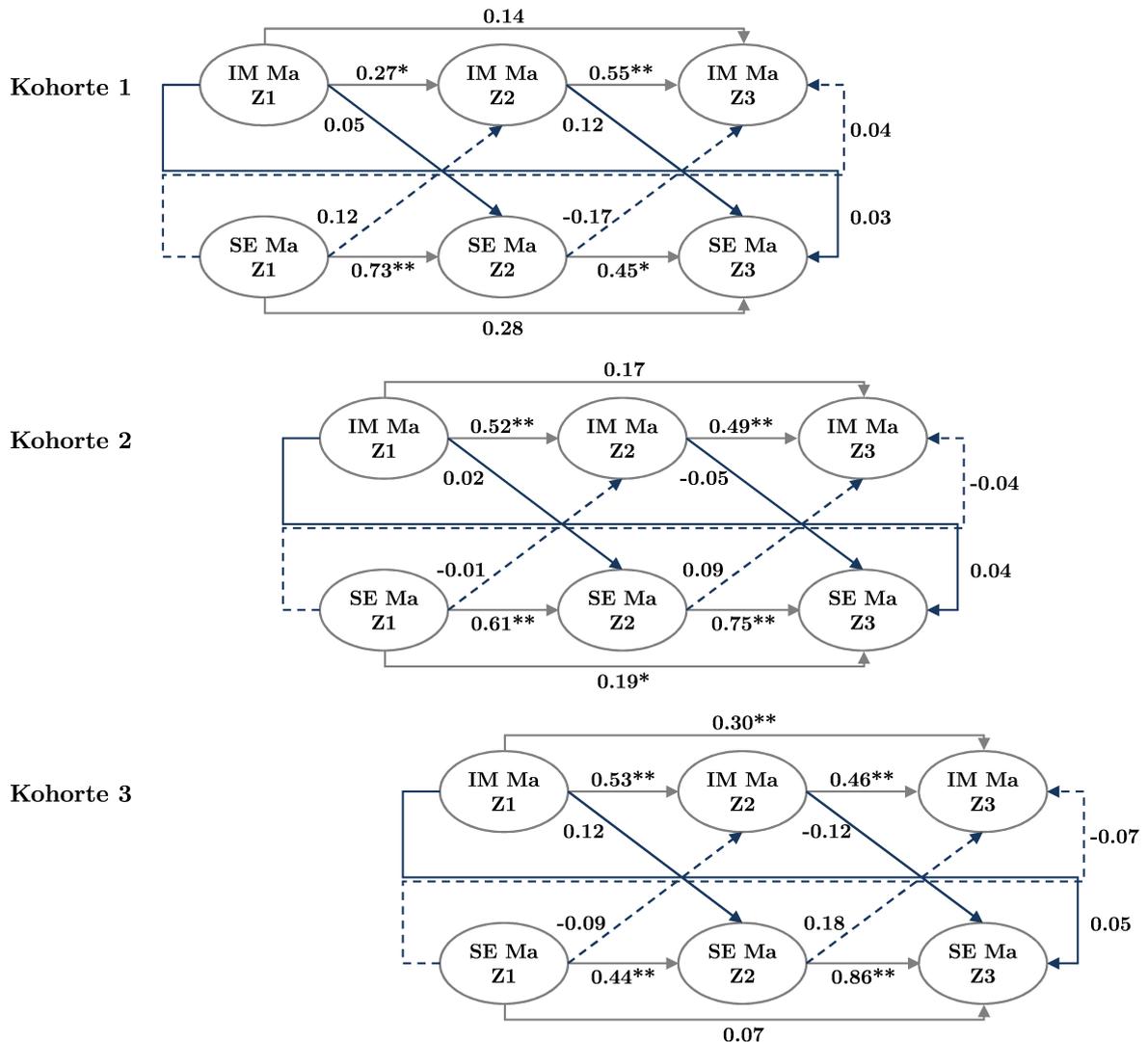


Abbildung 13.6. Multi-Kohorten Modell Intrinsische Motivation Mathematik. * $p < 0.05$; ** $p < 0.01$.

(von $\beta = -0.21$ auf $\beta = -0.14$ bzw. von $\beta = 0.16$ auf $\beta = 0.12$). In K1 blieb der Einfluss von SE auf FSK vergleichbar, allerdings stieg der Einfluss von FSK auf SE von $\beta = -0.03$ auf $\beta = 0.14$.

Tabelle 13.13 enthält schließlich die durchgeführte Modellprüfung im MKSD, die analog zu den Modellen ohne Kontrolle von g durchgeführt wurde. Insgesamt betrachtet lieferte diese Analyse das gleiche Resultat wie die zuvor durchgeführte Prüfung. Es zeigte sich analog ein signifikant schlechterer Modell-Fit bei Festsetzung der Pfade von SE auf FSK auf 0 (Modell 9). Die Prüfung der Höhe des Entwicklungseffektes ergab ein vergleichbares Ergebnis, allerdings fiel der Effekt nun etwas geringer aus verglichen mit dem Modell ohne Kontrolle von g (Modell 14).

Tabelle 13.13

Multigruppen Modelle Fähigkeitsselbstkonzept Mathematik mit Kontrolle von g

FSK_Ma	Modell	Vergleich	χ^2	df	CFI	TLI	SRMR	RMSEA	p	AIC	Δdf	$\Delta\chi^2(TRd)$
	M2		487.4**	359	0.97	0.96	0.08	0.04	0.88	23 593.15		
	M3	M2	507.4**	383	0.97	0.96	0.09	0.04	0.94	23 574.21	24	15.66
	M4	M3	545.84**	397	0.96	0.95	0.11	0.04	0.85	23 588.68	14	32.22**
	M5	M4	666.80**	433	0.94	0.93	0.13	0.05	0.25	23 646.12	36	108.66**
	M6	M5	699.59**	463	0.94	0.94	0.14	0.05	0.35	23 636.05	30	23.57
	M6a	M3	544.43**	413	0.97	0.96	0.09	0.04	0.95	23 561.02	30	29.45
	M7	M6a	549.09**	419	0.97	0.96	0.09	0.04	0.96	23 554.61	6	4.20
	M8	M7	551.71**	425	0.97	0.96	0.09	0.04	0.97	23 547.13	6	2.09
	M9	M8	573.74**	434	0.97	0.96	0.10	0.04	0.95	23 552.88	9	19.31*
	M10	M8	554.24**	434	0.97	0.97	0.09	0.04	0.99	23 533.47	9	2.13
FSK_Ma	M11		362.83**	272	0.97	0.96	0.10	0.04	0.83	14 731.28		
	M12	M11	364.78**	278	0.97	0.96	0.10	0.04	0.88	14 722.70	6	1.57
	M13	M12	372.85**	281	0.96	0.96	0.10	0.04	0.85	14 725.97	3	6.48
	M14	M12	365.26**	279	0.97	0.96	0.10	0.04	0.89	14 721.42	1	0.36
											3	6.12

Anmerkungen. CFI = Comparative Fit Index; TLI = Tucker-Lewis-Index; SRMR = Standardized Root Mean Square Residual; RMSEA = Root-Mean-Square-Error of Approximation; p = $P(RMSEA < 0.05)$; AIC = Akaike Information Criterion; Δdf = Differenz der Freiheitsgrade generierter Modelle; $\Delta\chi^2(TRd)$ = Satorra-Bentler Chi-Quadrat Differenztest; M2 = M1 der Full-Forward Modelle, alle Parameter zwischen den Kohorten frei; M3 = Faktorladungen zwischen den Kohorten gleichgesetzt; M4 = M3 und Faktorvarianzen zwischen den Kohorten; M5 = M4 und Intercepts zwischen den Kohorten; M6 = M5 und korrelierte Fehler zwischen den Kohorten; M6a = M4 und korrelierte Fehler zwischen den Kohorten; M7 = M6a und Stabilität FSK zwischen den Kohorten gleichgesetzt; M8 = M7 und Stabilität SE zwischen den Kohorten gleichgesetzt. M9 = M8 und Pfade von SE auf FSK auf 0 fixiert; M10 = M8 und Pfade von FSK auf SE auf 0 fixiert; M11 = M8 mit K1 und K3; M12 = M8 und Pfade von FSK auf SE auf 0 fixiert; M13 = M8 und Pfade von SE auf FSK innerhalb und zwischen den Kohorten gleichgesetzt; M14 = Abschätzung des Kohorteneffektes, Pfade im überlappenden Zeitfenster gleichgesetzt; Abnahme des Fits der aus dieser Restriktion resultiert, bildet den Kohorteneffekt ab. Verbleibende Abnahme im Fit von M12 zu M13 kann als Entwicklungseffekt interpretiert werden.

* $p < 0.05$.

** $p < 0.01$.

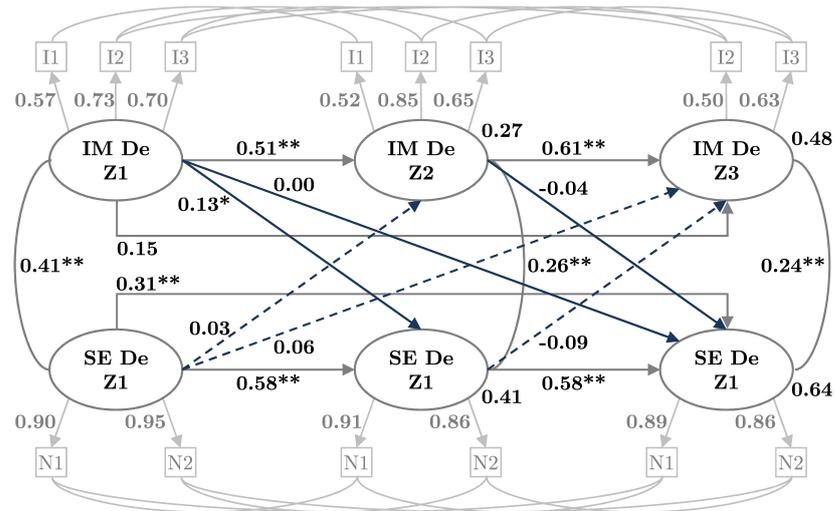


Abbildung 13.7. Full-Forward Modell Intrinsische Motivation Deutsch.
 * $p < 0.05$; ** $p < 0.01$.

Nach dieser Analyse der phänotypischen Zusammenhänge zwischen FSK, IM und SE unter Berücksichtigung der allgemeinen kognitiven Fähigkeit wurden auf Basis verhaltensgenetischer Modelle die Ursachen interindividueller Unterschiede in den Variablen selbst, sowie deren Stabilität untersucht.

2. Klasse 3. Klasse 4. Klasse 5. Klasse 6. Klasse 7. Klasse 8. Klasse

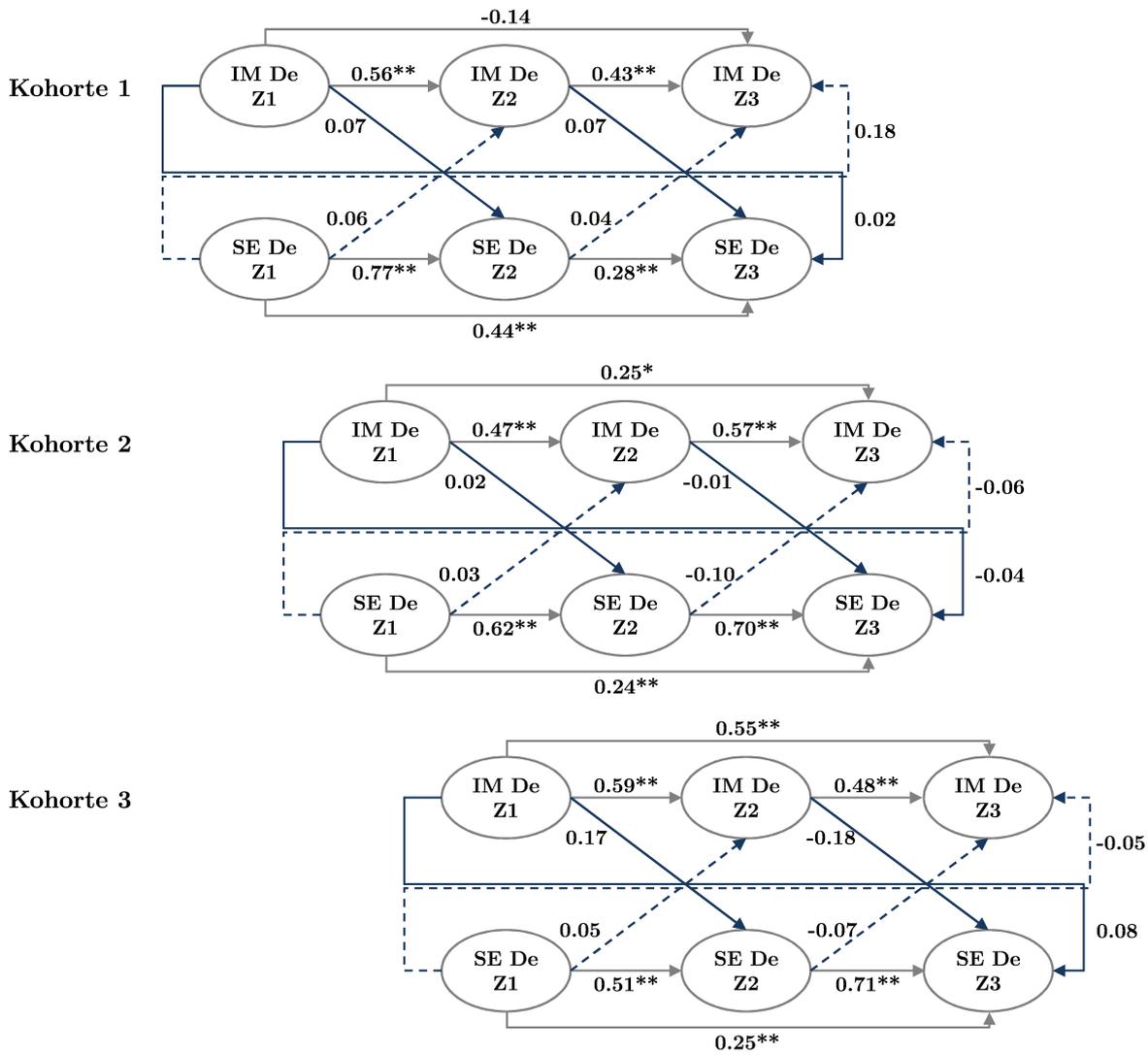


Abbildung 13.8. Multi-Kohorten Modell Intrinsische Motivation Deutsch.
 * $p < 0.05$; ** $p < 0.01$.

2. Klasse 3. Klasse 4. Klasse 5. Klasse 6. Klasse 7. Klasse 8. Klasse

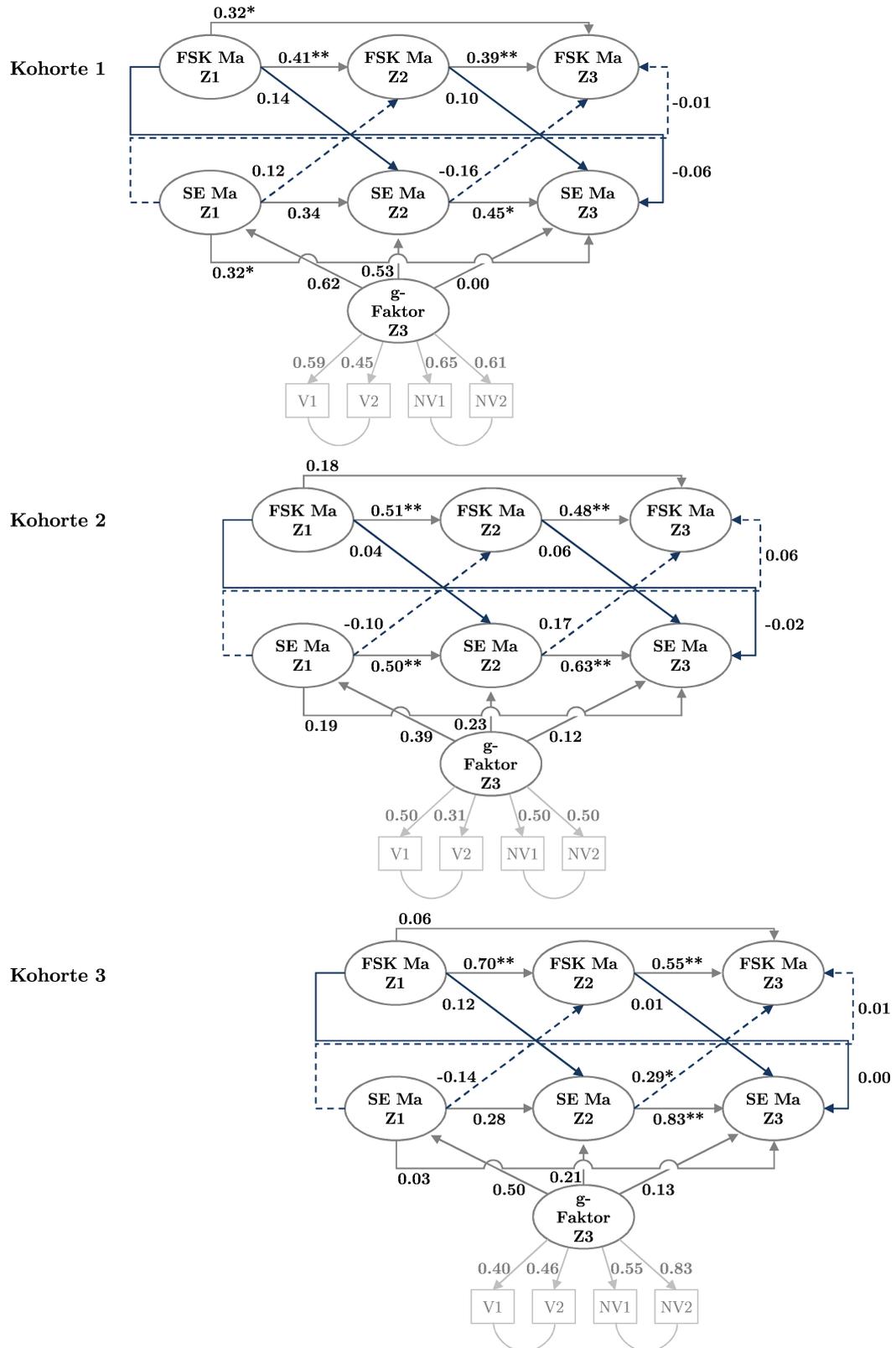


Abbildung 13.9. Multi-Kohorten Modell Fähigkeitsselbstkonzept Mathematik mit Kontrolle von g . * $p < 0.05$; ** $p < 0.01$.

14 Ergebnisse zur Beantwortung der ätiologischen Fragestellungen

Zur Beantwortung der unter Kapitel 6 formulierten Fragestellungen wurde auf univariate und multivariate genetische Analysen zurückgegriffen. Die hierfür zugrunde gelegte Stichprobe wurde in Kapitel 7.2.2 bereits eingehend beschrieben und auf die mit der Stichprobenauswahl verbundene eingeschränkte Vergleichbarkeit zwischen phänotypischen und ätiologischen Ergebnissen, die sich durch die fehlende Möglichkeit der kohortenspezifischen Auswertung ergab, hingewiesen. Die Analysen wurden, wie unter Kapitel 10 erläutert, mit alters- und geschlechtskorrigierten Daten durchgeführt.

14.1 Univariate verhaltensgenetische Analysen

In Tabelle 14.1 finden sich die ICCs der EZ und ZZ für die manifesten Skalen von SE, FSK, IM und dem g -Faktor zu den drei Messzeitpunkten. Erwartungsgemäß lagen die EZ Korrelationen jeweils höher als die ZZ Korrelationen, was auf einen genetischen Einfluss schließen ließ.

In SE Mathematik lag die Ähnlichkeit der EZ zu $Z1$ und $Z3$ in gleicher Höhe (0.71 bzw. 0.72), die Ähnlichkeit der ZZ stieg hingegen leicht an (von 0.44 auf 0.51). Zu $Z2$ lagen die Zwillingsähnlichkeiten für EZ und ZZ insgesamt niedriger als zu $Z1$ bzw. $Z3$ (EZ: 0.59; ZZ: 0.37). In SE Deutsch zeigte sich zu $Z2$ sowohl bei den EZ als auch bei den ZZ eine höhere Ähnlichkeit als zu $Z1$ und $Z3$. Auch hier war zu beobachten, dass die Ähnlichkeit der ZZ leicht anstieg, während diejenige der EZ etwas abnahm. Abgeleitet aus den ICC Schätzungen ($ICC_{ZZ} > \frac{1}{2} ICC_{EZ}$) konnte für SE in Mathematik und Deutsch ein ACE Modell als vollständiges Modell angenommen werden. Für den g -Faktor konnte ebenfalls ein ACE Modell als vollständiges Modell angenommen werden. Auch hier waren dementsprechend neben additiven genetischen Effekten, Effekte der geteilten und nichtgeteilten Umwelt zu erwarten. In den motivationalen Variablen FSK und IM zeigte sich kein einheitliches Muster. Hier waren nur im FSK Deutsch sowie der IM Deutsch zu $Z1$ auf Grundlage der gefundenen ICCs geteilte Umwelteffekte zu vermuten. In den übrigen Variablen lagen die $ICC_{ZZ} < \frac{1}{2} ICC_{EZ}$. Es wurden daher in der Folge ADE Modelle als vollständige Modelle für das FSK und die IM in Mathematik

sowie der IM Deutsch zu $Z2$ und $Z3$ geschätzt. Wie aus Tabelle 14.1 hervorgeht, stieg die Ähnlichkeit innerhalb der EZ Paare im FSK Mathematik, IM Mathematik und IM Deutsch dabei von $Z1$ nach $Z3$ an, während sich die ZZ eher unähnlicher wurden. Im FSK Deutsch sanken hingegen sowohl die Ähnlichkeiten der EZ als auch die der ZZ.

Tabelle 14.1

Intraklassenkorrelation der Variablen zu den verschiedenen Messzeitpunkten

	Zeitpunkt 1		Zeitpunkt 2		Zeitpunkt 3	
	EZ	ZZ	EZ	ZZ	EZ	ZZ
SE_Ma	0.71 [0.59; 0.80]	0.44 [0.30; 0.54]	0.59 [0.44; 0.71]	0.37 [0.24; 0.49]	0.72 [0.61; 0.81]	0.51 [0.39; 0.61]
SE_De	0.69 [0.57; 0.79]	0.42 [0.29; 0.53]	0.75 [0.65; 0.83]	0.53 [0.42; 0.63]	0.64 [0.50; 0.74]	0.50 [0.39; 0.61]
g-Faktor					0.66 [0.47; 0.77]	0.55 [0.40; 0.66]
FSK_Ma	0.40 [0.24; 0.53]	0.22 [0.10; 0.34]	0.43 [0.25; 0.58]	0.13 [-0.02; 0.27]	0.54 [0.19; 0.53]	0.02 [-0.16; 0.13]
FSK_De	0.55 [0.42; 0.66]	0.30 [0.19; 0.41]	0.43 [0.25; 0.58]	0.24 [0.10; 0.37]	0.40 [0.22; 0.55]	0.25 [0.10; 0.38]
IM_Ma	0.39 [0.24; 0.53]	0.17 [0.04; 0.29]	0.47 [0.30; 0.61]	0.08 [-0.07; 0.22]	0.46 [0.29; 0.60]	0.14 [-0.01; 0.28]
IM_De	0.31 [0.14; 0.45]	0.20 [0.08; 0.32]	0.58 [0.43; 0.70]	0.06 [0.06; 0.34]	0.52 [0.36; 0.65]	0.05 [-0.09; 0.20]

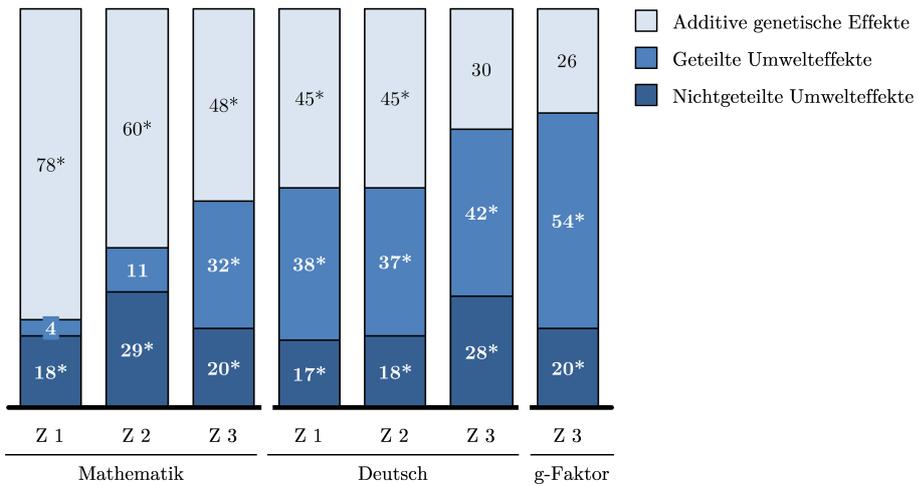
Anmerkungen. EZ = Eineiige Zwillinge; ZZ = Zweieiige Zwillinge; N(EZ) = 128 Paare ($Z1$), 94 Paare ($Z2$); N(ZZ) = 232 Paare ($Z1$), 173 Paare ($Z2$); SE = Schulerfolg; FSK = Fähigkeitsselbstkonzept; IM = Intrinsische Motivation; Ma = Mathematik; De = Deutsch; g-Faktor = Faktorwert der explorativen Faktorenanalyse; Konfidenzintervall (5%) in Klammern.

Um Äquivalenz zur latenten phänotypischen Modellierung herzustellen, wurden die univariaten und multivariaten Modelle als latente Modelle spezifiziert. Die Ergebnisse der Modellanpassung sind Tabellen 14.2, 14.3 sowie 14.4 zu entnehmen, die Parameterschätzungen der vollständigen Modelle sind in Abbildung 14.1 dargestellt.

Wie aus Abbildung 14.1 hervorgeht, ergab sich aus der Modellschätzung sowohl im SE Mathematik als im SE Deutsch quantitative Veränderungen in den Parameterschätzungen über den Untersuchungsverlauf hinweg. In beiden Fällen sank der Anteil genetischer Effekte mit zunehmendem Alter der Kinder, wobei dieser Trend vor allem in Mathematik zu beobachten war. Hier lag der genetische Effekt zu $Z1$ noch bei $a^2 = 0.78$, zu $Z3$ fand sich ein genetischer Effekt von $a^2 = 0.48$. Im Vergleich zu SE Deutsch lag dementsprechend der genetische Anteil insbesondere zu $Z1$ wesentlich höher, während dieser zu $Z3$ dann in etwa gleicher Höhe lag wie in SE Deutsch zu $Z1$ und $Z2$ ($a^2 = 0.45$). In Deutsch sank der genetische Anteil nach $Z3$ auf $a^2 = 0.30$. Der Einfluss der geteilten

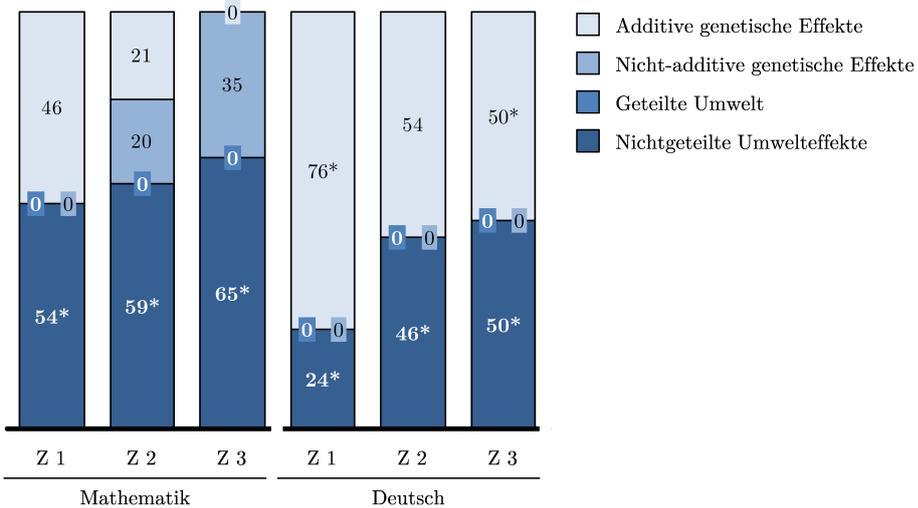
Schulerfolg und kognitive Fähigkeiten

Prozent erklärte Varianz



Fähigkeitsselbstkonzept

Prozent erklärte Varianz



Intrinsische Motivation

Prozent erklärte Varianz

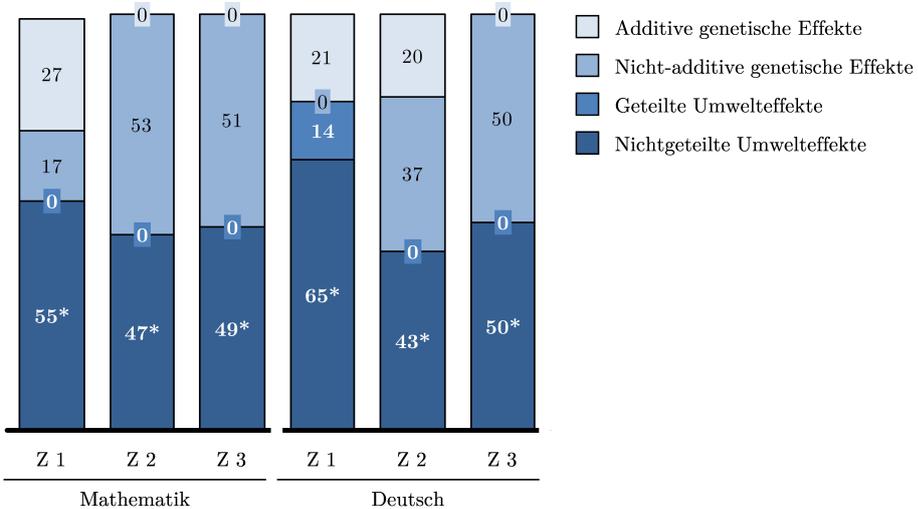


Abbildung 14.1. Parameterschätzung der latenten univariaten vollständigen Modelle. * $p < 0.05$.

Umwelt wurde in Mathematik hingegen deutlich größer ($c^2 = 0.04$ zu $Z1$, $c^2 = 0.32$ zu $Z3$), in Deutsch blieb dieser hingegen relativ konstant in den unterschiedlichen Altersklassen bestehen ($c^2 = 0.38, 0.37$ und 0.42). In Bezug auf den g -Faktor ließ sich feststellen, dass hier vor allem geteilte Umwelteinflüsse bedeutsam waren ($c^2 = 0.54$), während der Einfluss der nichtgeteilten Umwelt mit $e^2 = 0.20$ eher niedrig ausfiel. Für das FSK in Mathematik und Deutsch ergab die Parameterschätzung schließlich, dass die gesamte Varianz durch genetische Einflüsse (additive und nicht-additive) und Einflüsse der nichtgeteilten Umwelt erklärt werden konnte. Es zeigten sich hingegen keinerlei geteilte Umwelteinflüsse als bedeutsam bei der Erklärung interindividueller Unterschiede im FSK.

Wird die Veränderung in der Höhe der Parameterschätzung in den unterschiedlichen Altersbereichen betrachtet kann festgestellt werden, dass im FSK Mathematik der Anteil nichtgeteilter Umwelt relativ konstant bei ca. 60% bestehen blieb. Auf Seiten der genetischen Effekte zeigte sich in der Parameterschätzung eine Veränderung von additiven auf nicht-additive genetische Effekte. Für das FSK Deutsch verringerte sich der Anteil genetischer Effekte von $Z1$ nach $Z3$ (von 0.76 auf 0.50), hier fanden sich zudem ausschließlich additive genetische Effekte. Der Anteil an nichtgeteilten Umwelteffekten stieg entsprechend über die Zeit hinweg im FSK Deutsch an. Analog zum FSK Mathematik fand sich auch in der IM Mathematik keine wesentliche Veränderung in der Höhe genetischer Beeinflussung, wohl aber eine Veränderung der Schätzung von additiven zu nicht-additiven genetischen Effekten. Gleiches war auch für die IM Deutsch zu beobachten, hier lag allerdings zusätzlich eine Erhöhung des genetischen Anteils von $Z1$ nach $Z3$ vor ($a^2 = 0.21$ zu $Z1$ und $d^2 = 0.50$ zu $Z3$). Für IM Deutsch fand sich zu $Z1$ noch ein geringer Anteil geteilter Umwelteffekte ($c^2 = 0.14$), dieser war allerdings zu $Z2$ und $Z3$ nicht mehr zu beobachten.

Aus den durchgeführten Modellanpassungen wurden die jeweils best-fittenden, sparsamsten Modelle zur Erklärung interindividueller Unterschiede in den Variablen abgeleitet.

Für SE Mathematik (vgl. Tabelle 14.2) ergaben sich hier in allen Altersbereichen AE Modelle, die einen moderaten, bzw. gerade noch akzeptablen Modell-Fit aufwiesen. Für SE Deutsch fittete zu $Z1$ und $Z2$ jeweils das vollständige ACE Modell am besten, in $Z3$ war dies ein CE Modell. Hier war allerdings der Modell-Fit nur in $Z2$ mit einem RMSEA von 0.03 zufriedenstellend. Zu $Z1$ und $Z3$ wies der RMSEA von 0.13 bzw. 0.14 auf eine nur mäßige Modellpassung hin. Die hieraus resultierenden Parameterschätzungen können dementsprechend lediglich als Tendenzen interpretiert werden.

Für den latenten g -Faktor fittete ein CE Modell nicht signifikant schlechter als das voll-

Tabelle 14.2

Modell-Fit der latenten univariaten Modelle – Schulerfolg und g

		Modell	χ^2	df	p	AIC	RMSEA	$\Delta\chi^2$	Δ df	p
SE_Ma	Z1	ACE	25.17	7	0.00	11.17	0.12			
		AE	25.27	10	0.01	5.27	0.09	0.09	3	0.99
		CE	58.46	10	0.00	38.46	0.17	33.29	3	0.00
		E	166.05	13	0.00	140.05	0.27	140.88	6	0.00
	Z2	ACE	18.03	7	0.01	4.03	0.07			
		AE	19.23	10	0.04	-0.76	0.06	1.21	3	0.75
		CE	26.17	10	0.00	6.17	0.11	8.14	3	0.04
		E	90.86	13	0.00	64.86	0.22	72.84	6	0.00
	Z3	ACE	24.40	7	0.00	10.39	0.12			
		AE	30.14	10	0.00	10.14	0.09	5.74	3	0.13
		CE	37.96	10	0.00	17.96	0.15	13.57	3	0.00
		E	153.61	13	0.00	127.61	0.30	129.21	6	0.00
SE_De	Z1	ACE	27.79	7	0.00	13.79	0.13			
		AE	40.93	10	0.00	20.93	0.13	13.13	3	0.00
		CE	59.61	10	0.00	39.61	0.17	31.82	3	0.00
		E	297.02	13	0.00	271.02	0.37	269.23	6	0.00
	Z2	ACE	7.29	7	0.40	-6.71	0.03			
		AE	16.05	10	0.10	-3.95	0.05	8.76	3	0.03
		CE	18.02	10	0.06	-1.98	0.08	10.73	3	0.01
		E	153.78	13	0.00	127.78	0.30	146.49	6	0.00
	Z3	ACE	20.62	7	0.00	6.62	0.12			
		AE	28.22	10	0.00	8.22	0.12	7.60	3	0.06
		CE	26.76	10	0.00	6.76	0.10	6.14	3	0.11
		E	124.87	13	0.00	98.87	0.27	104.24	6	0.00
g-Faktor	Z3	ACE	175.60	49	0.00	77.60	0.14			
		AE	189.82	54	0.00	81.82	0.14	14.22	5	0.01
		CE	182.89	54	0.00	74.89	0.13	7.29	5	0.20
		E	359.08	59	0.00	241.08	0.20	183.48	10	0.00

Anmerkungen. Anzahl Paare in der Analyse: 360 (Z1), 267 (Z2 und Z3); SE=Schulerfolg; Ma=Mathematik; De=Deutsch; ACE: Modell welches additive genetische (A), geteilte (C) und nichtgeteilte Umwelteinflüsse (E) annimmt; ADE: Modell welches additive genetische (A), nicht-additive genetische (D) und nichtgeteilte Umwelteinflüsse (E) annimmt; CE: Modell welches geteilte (C) und nichtgeteilte Umwelteinflüsse (E) annimmt; DE: Modell welches nicht-additive genetische (D) und nichtgeteilte Umwelteinflüsse (E) annimmt; E: Modell welches nichtgeteilte Umwelteinflüsse (E) annimmt; p=Signifikanzlevel des χ^2 bei Anzahl df; AIC= Akaike Information Criterion; RMSEA=Root-Mean-Square-Error of Approximation; $\Delta\chi^2$ =Likelihood Ratio χ^2 -Test mit Δ df; Best-fittendes Modell *fett* markiert.

ständige ACE Modell. Allerdings wurde auch hier insgesamt kein adäquater Modell-Fit erreicht ($\chi^2 = 182.89$, $df = 54$, $p = 0.00$, AIC = 74.89, RMSEA = 0.13).

Der Fit der Modelle für die Motivationsvariablen war hingegen durchweg zufriedenstellend bis gut (vgl. Tabellen 14.3 und 14.4). Im FSK Mathematik ergab sich, wie bereits bei Betrachtung der Parameterschätzungen in Abbildung 14.1 zu vermuten war, ein

DE Modell als bestes und sparsamstes Modell. Da jedoch die Annahme nicht-additiver genetischer Effekte ohne additive genetische Effekte wenig plausibel erscheint (vgl. Kapitel 10) wurde hier jeweils das vollständige Modell (ADE) in den Folgeanalysen herangezogen. Gleiches galt für die IM Deutsch und Mathematik zu Z2 und Z3. Im FSK Deutsch sowie der IM Deutsch und Mathematik zu Z1 erwies sich ein AE Modell in allen Altersbereichen als nicht signifikant schlechter als das vollständige Modell.

Tabelle 14.3
Modell-Fit der latenten univariaten Modelle – Fähigkeitsselbstkonzept

	Modell	χ^2	df	p	AIC	RMSEA	$\Delta\chi^2$	Δ df	p	
FSK_Ma	Z1	ADE	29.05	24	0.22	-18.95	0.02			
		AE	32.16	28	0.27	-23.84	0.03	3.12	4	0.54
		DE	33.92	28	0.20	-22.09	0.03	4.89	4	0.30
		E	83.81	32	0.00	19.81	0.10	54.76	8	0.00
	Z2	ADE	32.15	24	0.12	-15.85	0.04			
		AE	34.12	28	0.20	-21.88	0.04	1.97	4	0.74
		DE	32.47	28	0.26	-23.53	0.04	0.32	4	0.99
		E	66.74	32	0.00	2.74	0.08	34.59	8	0.00
	Z3	ADE	44.73	24	0.01	-3.27	0.07			
		AE	53.89	28	0.00	-1.86	0.08	9.16	4	0.06
		DE	44.73	28	0.02	-11.27	0.06	0.00	4	1.00
		E	70.70	32	0.00	6.70	0.10	25.97	8	0.00
FSK_De	Z1	ACE	53.54	24	0.00	5.54	0.08			
		AE	53.61	28	0.00	-2.39	0.07	0.07	4	1.00
		CE	73.52	28	0.00	17.52	0.10	19.98	4	0.00
		E	138.06	32	0.00	74.06	0.14	84.52	8	0.00
	Z2	ACE	47.18	27	0.00	-0.83	0.09			
		AE	47.18	28	0.01	-8.83	0.08	0.00	4	1.00
		CE	60.99	28	0.00	4.99	0.10	13.82	4	0.01
		E	105.62	32	0.00	41.62	0.14	58.44	8	0.00
	Z3	ACE	32.62	24	0.11	-15.39	0.05			
		AE	32.62	28	0.25	-23.39	0.03	0.00	4	1.00
		CE	48.21	28	0.01	-7.79	0.08	15.59	4	0.00
		E	92.37	32	0.00	28.37	0.12	59.76	8	0.00

Anmerkungen. Anzahl Paare in der Analyse: 360 (Z1), 267 (Z2 und Z3); FSK = Fähigkeitsselbstkonzept; Ma = Mathematik; De = Deutsch; ACE: Modell welches additive genetische (A), geteilte (C) und nichtgeteilte Umwelteinflüsse (E) annimmt; ADE: Modell welches additive genetische (A), nicht-additive genetische (D) und nichtgeteilte Umwelteinflüsse (E) annimmt; CE: Modell welches geteilte (C) und nichtgeteilte Umwelteinflüsse (E) annimmt; DE: Modell welches nicht-additive genetische (D) und nichtgeteilte Umwelteinflüsse (E) annimmt; E: Modell welches nichtgeteilte Umwelteinflüsse (E) annimmt; p = Signifikanzlevel des χ^2 bei Anzahl df; AIC = Akaike Information Criterion; *RMSEA* = Root-Mean-Square-Error of Approximation; $\Delta\chi^2$ = Likelihood Ratio χ^2 -Test mit Δ df; Best-fittendes Modell *fett* markiert.

Die resultierenden Parameterschätzungen der best-fittenden Modelle sind nachfolgend

Tabelle 14.4

Modell-Fit der latenten univariaten Modelle – Intrinsische Motivation

		Modell	χ^2	df	p	AIC	RMSEA	$\Delta\chi^2$	Δ df	p
IM_Ma	Z1	ADE	28.40	24	0.24	-19.60	0.03			
		AE	29.31	28	0.40	-26.69	0.03	0.91	4	0.92
		DE	59.85	28	0.37	-26.15	0.02	1.45	4	0.84
		E	70.49	32	0.00	6.49	0.08	42.09	8	0.00
	Z2	ADE	53.78	24	0.00	5.78	0.09			
		AE	62.43	28	0.00	6.42	0.10	8.65	4	0.07
		DE	53.78	28	0.00	-2.22	0.08	0.00	4	1.00
		E	92.46	32	0.00	28.46	0.12	38.68	8	0.00
	Z3	ADE	36.84	24	0.05	-11.16	0.06			
		AE	40.60	28	0.06	-15.40	0.06	3.76	4	0.44
		DE	36.95	28	0.12	-19.05	0.04	0.11	4	0.99
		E	72.38	32	0.00	8.38	0.09	35.54	8	0.00
IM_De	Z1	ACE	50.13	24	0.00	2.13	0.08			
		AE	50.80	28	0.01	-5.20	0.07	0.67	4	0.96
		CE	60.19	28	0.00	4.19	0.08	10.06	4	0.04
		E	95.63	32	0.00	31.63	0.11	45.50	8	0.00
	Z2	ADE	47.62	24	0.00	-0.38	0.09			
		AE	51.33	28	0.01	-4.67	0.08	3.72	4	0.45
		DE	48.14	28	0.01	-7.86	0.08	0.52	4	0.97
		E	128.69	32	0.00	64.69	0.16	81.08	8	0.00
	Z3	ADE	48.90	24	0.00	0.90	0.09			
		AE	56.84	28	0.00	0.84	0.09	7.94	4	0.09
		DE	48.90	28	0.01	-7.09	0.08	0.00	4	1.00
		E	98.94	32	0.00	34.94	0.12	50.04	8	0.00

Anmerkungen. Anzahl Paare in der Analyse: 360 (Z1), 267 (Z2 und Z3); IM = Intrinsische Motivation; Ma = Mathematik; De = Deutsch; ACE: Modell welches additive genetische (A), geteilte (C) und nichtgeteilte Umwelteinflüsse (E) annimmt; ADE: Modell welches additive genetische (A), nicht-additive genetische (D) und nichtgeteilte Umwelteinflüsse (E) annimmt; CE: Modell welches geteilte (C) und nichtgeteilte Umwelteinflüsse (E) annimmt; DE: Modell welches nicht-additive genetische (D) und nichtgeteilte Umwelteinflüsse (E) annimmt; E: Modell welches nichtgeteilte Umwelteinflüsse (E) annimmt; p = Signifikanzlevel des χ^2 bei Anzahl df; AIC = Akaike Information Criterion; RMSEA = Root-Mean-Square-Error of Approximation; $\Delta\chi^2$ = Likelihood Ratio χ^2 -Test mit Δ df; Best-fittendes Modell *fett* markiert.

in Tabelle 14.5 zusammengefasst. Wie aus Tabelle 14.5 hervorgeht, ergab die Schätzung im latenten Modell insbesondere in SE Mathematik sowie im FSK Deutsch zu Z1 einen hohen genetischen Effekt ($a^2 = 0.72 - 0.82$). Im g-Faktor ergab sich kein bedeutsamer genetischer Effekt. In den verbleibenden Skalen lag jeweils eine moderate genetische Beeinflussung vor.

Tabelle 14.5

Parameterschätzung der best-fittenden latenten univariaten Modelle

		a^2	c^2/d^2	e^2
SE_Ma	Z1	0.82 [0.74; 0.88]		0.18 [0.12; 0.26]
	Z2	0.72 [0.58; 0.82]		0.28 [0.18; 0.42]
	Z3	0.82 [0.73; 0.89]		0.18 [0.11; 0.27]
SE_De	Z1	0.45 [0.27; 0.65]	0.38 [0.19; 0.53]	0.17 [0.13; 0.24]
	Z2	0.45 [0.16; 0.75]	0.37 [0.10; 0.60]	0.18 [0.09; 0.29]
	Z3		0.62 [0.52; 0.71]	0.38 [0.29; 0.48]
g-Faktor	Z3		0.71 [0.59; 0.81]	0.29 [0.19; 0.41]
FSK_Ma	Z1	0.47 [0.29; 0.62]		0.53 [0.38; 0.71]
	Z2		0.43 [0.24; 0.58]	0.57 [0.42; 0.76]
	Z3		0.35 [0.12; 0.53]	0.65 [0.47; 0.88]
FSK_De	Z1	0.76 [0.60; 0.89]		0.24 [0.11; 0.40]
	Z2	0.54 [0.31; 0.72]		0.46 [0.28; 0.69]
	Z3	0.50 [0.27; 0.68]		0.50 [0.31; 0.73]
IM_Ma	Z1	0.43 [0.25; 0.59]		0.57 [0.41; 0.75]
	Z2		0.53 [0.32; 0.69]	0.47 [0.31; 0.67]
	Z3		0.51 [0.31; 0.66]	0.49 [0.34; 0.69]
IM_De	Z1	0.39 [0.19; 0.57]		0.61 [0.43; 0.81]
	Z2		0.58 [0.38; 0.74]	0.42 [0.26; 0.62]
	Z3		0.50 [0.02; 0.69]	0.50 [0.31; 0.73]

Anmerkungen. Anzahl Paare in der Analyse: 360 (Z1), 267 (Z2 und Z3); EZ = Eineiige Zwillinge; ZZ = Zweieiige Zwillinge; SE = Schulerfolg; FSK = Fähigkeitsselbstkonzept; IM = Intrinsische Motivation; De = Deutsch; Ma = Mathematik; a^2 = additive genetische Effekte; d^2 = nicht-additive genetische Effekte; c^2 = geteilte Umwelteffekte; e^2 = nichtgeteilte Umwelteffekte. KI (5%) in Klammern.

14.2 Multivariate verhaltensgenetische Analysen

In Bezug auf die multivariaten verhaltensgenetischen Fragestellungen interessierte vor allem der Übergang in die weiterführende Schule. Aus diesem Grund wurden, wie bereits in Kapitel 7.2.2 beschrieben, für die multivariaten genetischen Analysen die Daten von K2 und K3 zusammengefasst, da für beide Kohorten der Wechsel zum gleichen Zeitpunkt innerhalb der Untersuchung stattfand (zwischen $Z1$ und $Z2$). Bedingt durch die unterschiedlichen Stichproben der multivariaten im Vergleich zu den univariaten Analysen ergaben sich Abweichungen in den Parameterschätzungen. Aufgrund der recht geringen Stichprobengröße in Relation zur Komplexität der zu schätzenden Modelle kann es zudem zu instabilen Schätzungen bzw. großen Konfidenzintervallen kommen (vgl. Kapitel 10). Hierauf wird an entsprechender Stelle nochmals hingewiesen.

14.2.1 Ursachen interindividueller Stabilität

Als Voraussetzung einer längsschnittlichen genetischen Analyse einer Variable muss zunächst phänotypische Stabilität bestätigt werden. Liegt diese nicht vor, sind die genetischen und Umwelteinflüsse zu einem bestimmten Alter unabhängig von denjenigen zu einem anderen Alter (Kovas et al., 2007). Auf Grundlage der phänotypisch gefundenen mittleren bis hohen Stabilitäten der Variablen über einen Zeitraum von sechs Jahren, konnten unter Verwendung der in Kapitel 10.2 erläuterten Cholesky Zerlegungen, die Ursachen von Stabilität und Veränderung in SE, FSK und IM analysiert werden.¹

Die Ergebnisse der Cholesky Zerlegung können den Abbildungen 14.1 sowie 14.2 für den schulischen Erfolg, 14.3 und 14.5 für das FSK und schließlich 14.4 sowie 14.6 entnommen werden. Die Zerlegung erfolgte jeweils mit dem vollständigen Modell der univariaten Analyse, um ein möglichst umfassendes Bild der Zusammenhänge zu erhalten. Für SE Mathematik und Deutsch sowie FSK und IM Deutsch wurde entsprechend ein ACE, für FSK und IM Mathematik ein ADE Modell gewählt. In der folgenden Darstellung der Ergebnisse erfolgt die Betrachtung der Skalen in Mathematik und Deutsch zunächst für die genetischen Anteile hiernach werden die Umweltanteile berichtet. In Tabelle 14.6 ist der Modell-Fit der vollständigen Cholesky Modelle aufgeführt.

Tabelle 14.6
Modell-Fit der latenten multivariaten Modelle

	Modell	χ^2	df	p	AIC	RMSEA
SE_Ma	ACE	195.27	114	0.00	-32.73	0.09
SE_De	ACE	183.96	114	0.00	-44.04	0.09
FSK_Ma	ADE	542.52	288	0.00	-33.48	0.10
FSK_De	ACE	746.96	288	0.00	170.96	0.13
IM_Ma	ADE	463.99	288	0.00	-112.01	0.08
IM_De	ACE	753.34	288	0.00	177.34	0.13

Anmerkungen. Anzahl Paare in der Analyse: 63 EZ, 125 ZZ; SE = Schulerfolg; FSK = Fähigkeitsselbstkonzept; IM = Intrinsische Motivation; Ma = Mathematik; De = Deutsch; ACE: Modell welches additive genetische (A), geteilte (C) und nichtgeteilte Umwelteinflüsse (E) annimmt; ADE: Modell welches additive genetische (A), nicht-additive genetische (D) und nichtgeteilte Umwelteinflüsse (E) annimmt; p = Signifikanzlevel des χ^2 bei Anzahl df; AIC = Akaike Information Criterion; RMSEA = Root-Mean-Square-Error of Approximation; $\Delta\chi^2$ = Likelihood Ratio χ^2 -Test mit Δ df.

Schulerfolg Mathematik und Deutsch

Für SE in Mathematik und Deutsch konnte ein zufriedenstellender Fit erreicht werden, wenngleich der RMSEA in beiden Fällen mit 0.09 recht hoch ausfiel. Angesichts der Komplexität des zu schätzenden Modells kann dieser allerdings noch als ausreichend angenommen werden.²

Die signifikanten und substanziellen Ladungen des SE in Mathematik zu Z2 und Z3 auf den ersten genetischen Faktor (A_1) wiesen auf eine starke genetische Kontinuität hin. Von der genetischen Varianz zu Z2 ($0.33 + 0.15 = 0.48$) wurden ca. 69% ($0.33 \div 0.48$) durch gemeinsame genetische Effekte mit Z1 erklärt. Zu Z3 wurden 65% durch die genetische Varianz zu Z1 erklärt. Zu Z2 lag ein spezifischer genetischer Anteil in Höhe von 31% der gesamten genetischen Varianz (Z2) vor, wohingegen zu Z3 kein spezifischer genetischer Anteil festgestellt werden konnte. Die verbleibenden 35% der genetischen Varianz zu Z3, die nicht mit der genetischen Varianz zu Z1 geteilt waren, wurden mit der genetischen Varianz zu Z2 geteilt. Diese genetischen Einflüsse sind also als unabhängig von den genetischen Einflüssen zu sehen, die zu allen drei Zeitpunkten vorlagen. Zusammenfassend betrachtet konnte ein starker genetischer Faktor zu Z1 bestätigt werden. Dieser erklärte einen Großteil der genetischen Varianz zu Z2 und

¹ Die allgemeinen kognitiven Fähigkeiten werden aufgrund der fehlgeschlagenen latenten Modellierung der ersten kognitiven Messung an dieser Stelle nicht weiter betrachtet.

² Der RMSEA berücksichtigt die Stichprobengröße und belohnt sparsame Modelle (vgl. Kapitel 9.2).

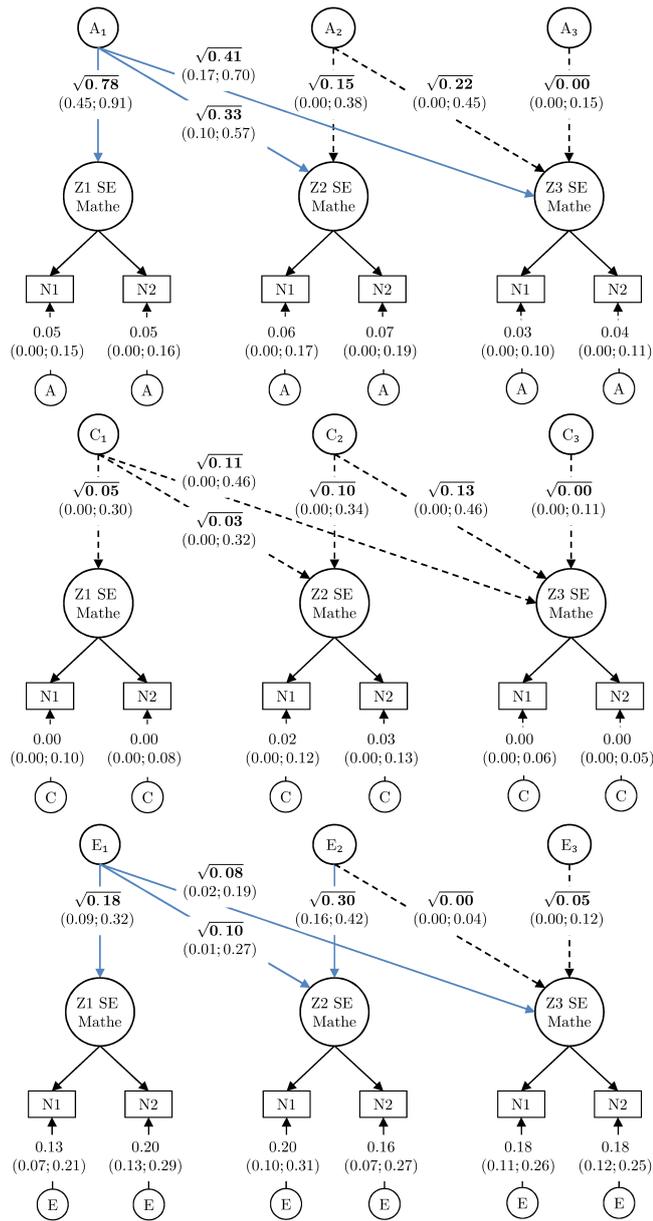


Abbildung 14.1. Cholesky Zerlegung des schulischen Erfolgs in Mathematik. Die Pfadschätzungen sind aus Gründen einer einfacheren Interpretierbarkeit als Wurzel des standardisierten partiellen Regressionsgewichtes dargestellt. Die Angabe in Klammern unter der Pfadschätzung gibt das 5% Konfidenzintervall der Schätzung an.

Z3. Zusätzlich ließen sich für Z2 noch spezifische genetische Effekte feststellen, die wiederum einen substantiellen Anteil der genetischen Varianz in Z3 erklärte. Zu Z3 fanden sich hingegen keine spezifischen genetischen Effekte mehr.

Ein ähnliches Befundmuster zeigte sich für den schulischen Erfolg in Deutsch. Auch hier lag eine hohe genetische Kontinuität vor. 60% der genetischen Varianz zu Z2 und 45% der genetischen Varianz zu Z3 wurden mit der genetischen Varianz zu Z1 geteilt. Im Unterschied zum Fach Mathematik fanden sich sowohl zu Z2 als auch zu Z3 spezifische

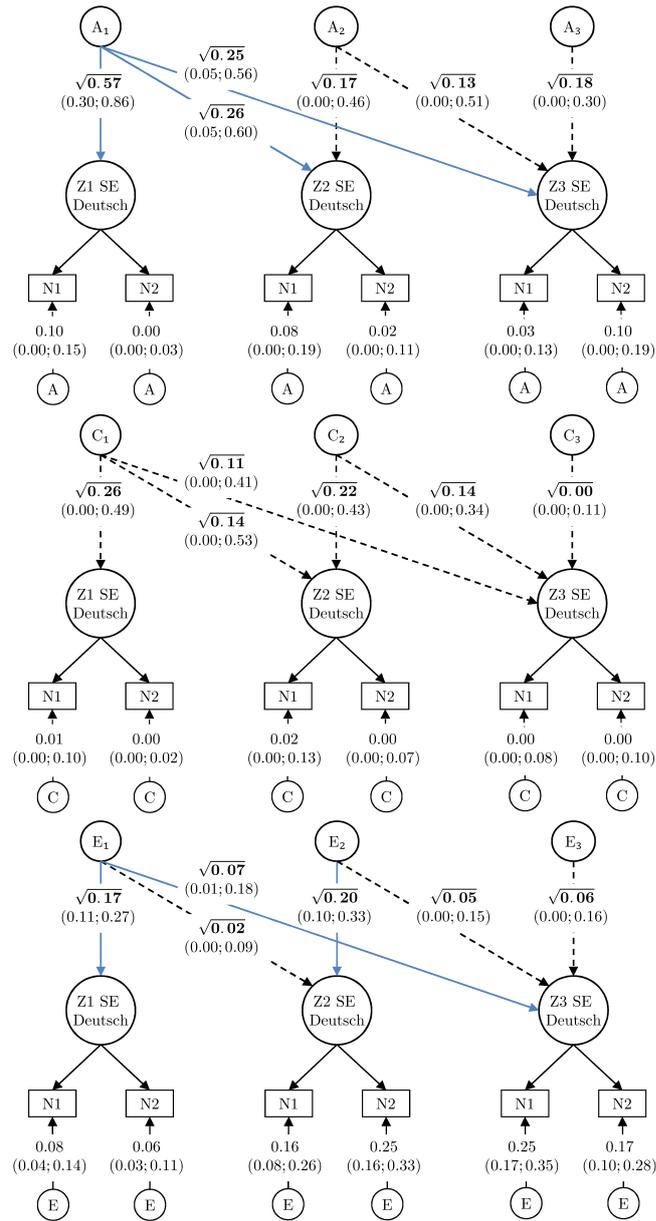


Abbildung 14.2. Cholesky Zerlegung des schulischen Erfolgs in Deutsch. Die Pfadschätzungen sind aus Gründen einer einfacheren Interpretierbarkeit als Wurzel des standardisierten partiellen Regressionsgewichtes dargestellt. Die Angabe in Klammern unter der Pfadschätzung gibt das 5% Konfidenzintervall der Schätzung an.

genetische Effekte, diese waren jedoch nicht signifikant. Schließlich fand sich auch im SE Deutsch eine genetische Kontinuität zwischen $Z2$ und $Z3$ dahingehend, dass 23% der genetischen Varianz zu $Z3$ mit der genetischen Varianz zu $Z2$ überlappte.

Die Ergebnisse für die geteilte Umwelt (C_1, C_2 und C_3) in Abbildung 14.1 und 14.2 können analog interpretiert werden, allerdings muss hier berücksichtigt werden, dass die Effekte der geteilten Umwelt generell sehr niedrig ausfielen (vgl. auch Tabelle 14.5).

Aufgrund dieser geringen Varianz der geteilten Umwelt ergaben sich bei der Zerlegung der Kovarianz zwischen den Altersstufen große Konfidenzintervalle. Für Mathematik ließ sich konstatieren, dass die Varianz der geteilten Umwelt insgesamt niedriger ausfiel als die genetische Varianz. Über den Untersuchungsverlauf hinweg stieg die geteilte Umweltvarianz von $c^2 = 0.05$ zu $Z1$ auf $c^2 = 0.13$ zu $Z2$ und schließlich $c^2 = 0.24$ zu $Z3$ an. Es zeigte sich außerdem, dass ein Großteil der geteilten Umweltvarianz zu $Z3$ (0.24) mit der Varianz zu $Z1$ (46%) und $Z2$ (51%) überlappte.

Im SE Deutsch fand sich hingegen ein höheres Maß an geteilter Umweltvarianz. Für diese zeigte sich das gleiche Muster wie für die genetische Varianz: Eine recht hohe Überlappung zwischen den Zeitpunkten, ein spezifischer Anteil zu $Z2$ und keinerlei spezifische geteilte Umweltvarianz zu $Z3$ mehr.

Für die nichtgeteilte Umweltvarianz ergab die Cholesky Zerlegung für SE in Mathematik und Deutsch übereinstimmend vor allem messzeitpunktspezifische Varianzanteile. Diese waren insbesondere für den zweiten Erhebungszeitpunkt zu beobachten, während zu $Z3$ lediglich 13% (SE Mathematik) bzw. 18% (SE Deutsch) der phänotypischen Varianz auf nichtgeteilte Umwelteffekte zurückgingen.

Neben dieser Betrachtung der Varianz zu den unterschiedlichen Messzeitpunkten und deren Überlappung zwischen den Messzeitpunkten wurde die bivariate Erblichkeit bzw. die bivariate geteilte und nichtgeteilte Umwelt betrachtet. Wie in Kapitel 10 erläutert, gibt die bivariate Erblichkeit das Ausmaß an, zudem die phänotypische Korrelation zwischen den Messzeitpunkten (also die phänotypische Stabilität) genetisch vermittelt ist, wobei dies unabhängig von der absoluten Höhe der Korrelation ist. Wie Tabelle 14.7 zu entnehmen ist, lag die bivariate Erblichkeit für SE Mathematik zwischen 0.68 und 0.75 und zwischen 0.48 und 0.61 für SE Deutsch. Dies bedeutet, dass beispielsweise in SE Mathematik ca. zwei drittel der Stabilität auf genetische Einflüsse zurückzuführen waren. Die bivariate geteilte Umwelt lag in beiden Fällen mit Werten zwischen 0.06 und 0.36 weitaus niedriger. Schließlich wurde die phänotypische Korrelation zwischen 9% und 20% durch überlappende nichtgeteilte Umwelteffekte vermittelt.

FSK und IM Mathematik

Für FSK und IM Mathematik wurden jeweils multivariate ADE Modelle geschätzt. Wie aus Tabelle 14.6 hervorgeht, lag auch für diese Variablen der Fit in einem noch akzeptablen Bereich, wenn die Stichprobengröße und die Komplexität des Modells berücksichtigt werden (RMSEA = 0.10 bzw. RMSEA = 0.08). Auch hier fielen die Konfidenzintervalle aufgrund der geringen Power der Stichprobe sehr groß aus. Werden die Ergebnisse der Cholesky Zerlegung betrachtet (vgl. Abbildung 14.3) ist zunächst ersichtlich, dass vor allem nicht-additive genetische Effekte vorlagen. Im FSK lag der additive genetische Effekt zu $Z1$ mit 0.22 am höchsten, zu $Z2$ lag dieser bei 0.16 und

Tabelle 14.7
Bivariate Erblichkeit und bivariate geteilte und nichtgeteilte Umwelt

		Z1 - Z2	Z2 - Z3	Z1 - Z3
SE_Ma	biv _{a2}	0.75	0.74	0.68
	biv _{c2}	0.05	0.10	0.21
	biv _{e2}	0.20	0.16	0.11
SE_De	biv _{a2}	0.61	0.58	0.48
	biv _{c2}	0.30	0.26	0.36
	biv _{e2}	0.09	0.17	0.16
FSK_Ma	biv _{a2}	0.28	0.08	0.04
	biv _{d2}	0.43	0.53	0.42
	biv _{e2}	0.29	0.39	0.53
FSK_De	biv _{a2}	0.93	0.85	0.70
	biv _{c2}	0.00	0.00	0.00
	biv _{e2}	0.07	0.15	0.30
IM_Ma	biv _{a2}	0.06	0.16	0.11
	biv _{d2}	0.64	0.66	0.44
	biv _{e2}	0.29	0.18	0.46
IM_De	biv _{a2}	0.82	0.64	0.68
	biv _{c2}	0.00	0.00	0.00
	biv _{e2}	0.18	0.36	0.32

Anmerkungen. Anzahl Paare in der Analyse: 63 EZ, 125 ZZ; SE = Schulerfolg; FSK = Fähigkeitsselbstkonzept; IM = Intrinsische Motivation; Ma = Mathematik; De = Deutsch; biv_{a2} = bivariate Erblichkeit; biv_{c2} = bivariate geteilte Umwelt; biv_{e2} = bivariate nichtgeteilte Umwelt; biv_{d2} = bivariate nicht-additive Erblichkeit.

zu Z3 bei 0.01. Die additiven genetischen Effekte zu Z2 überlappten zu ca. 69% mit denen zu Z1. Bei der IM Mathematik lag einzig zum dritten Zeitpunkt ein additive genetische Beeinflussung vor ($a^2 = 0.26$), diese überlappte zu 80% mit den genetischen Effekten zu Z2. Für die nicht-additiven genetischen Effekte zeigten sich im FSK hohe Ladungen auf den ersten genetischen Faktor: Von der genetischen Varianz zu Z2 wurden ca. 82% und von der zu Z3 zu 97% geteilt. Bei der IM Mathematik fand sich neben einem starken nicht-additiven genetischen Faktor zu Z1, der zu ca. 50% mit dem nicht-additiven genetischen Faktor zu Z2 und zu 100% zu Z3 überlappte, zusätzlich spezifische nicht-additive genetische Effekte zu Z2 (0.31). Für die nichtgeteilte Umwelt ergab sich für das FSK und die IM in Mathematik ein identisches Muster: Hier waren die Effekte zu Z1 und Z2 in erster Linie messzeitpunktspezifisch und auch zu Z3 war ein substanzieller Anteil der Effekte spezifisch für diesen Messzeitpunkt. Es zeigte sich jedoch auch, dass 50% (FSK) bzw. 41% (IM) der Varianz in nichtgeteilten Umweltef-

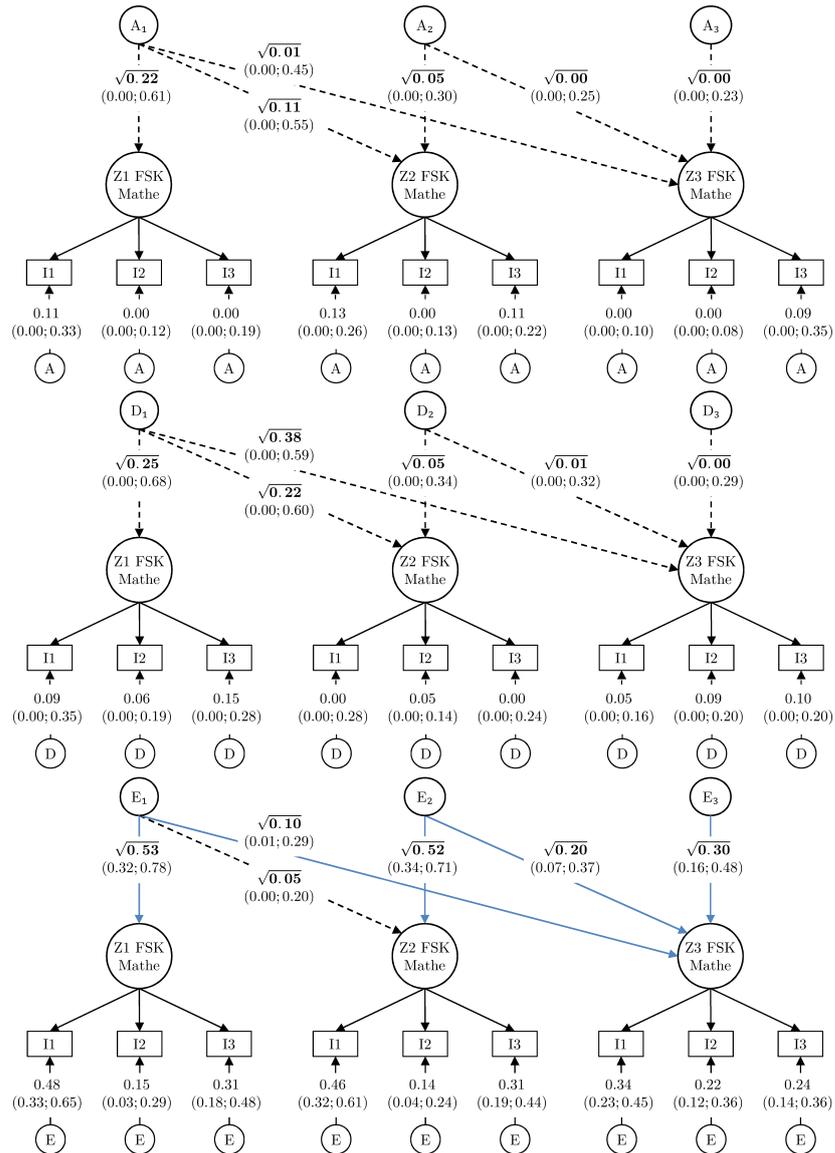


Abbildung 14.3. Cholesky Zerlegung des Fähigkeitsselbstkonzeptes Mathematik. Die Pfadschätzungen sind aus Gründen einer einfacheren Interpretierbarkeit als Wurzel des standardisierten partiellen Regressionsgewichtes dargestellt. Die Angabe in Klammern unter der Pfadschätzung gibt das 5% Konfidenzintervall der Schätzung an.

fekten zu Z3 mit denen zu Z2 geteilt wurden.

Für die bivariate Erbllichkeit bzw. die bivariate nichtgeteilte Umwelt ergab sich in beiden Motivationsvariablen ein ähnliches Muster, der Großteil der phänotypischen Korrelation zwischen den verschiedenen Zeitpunkten war jeweils durch genetische Effekte, insbesondere nicht-additive, vermittelt (zwischen 0.04 und 0.28 für biv_{a_2} und zwischen 0.42 und 0.66 für $biv_{a_2^2}$). Zwischen 18% und 53% der phänotypischen Korrelationen waren durch die nichtgeteilte Umwelt vermittelt.

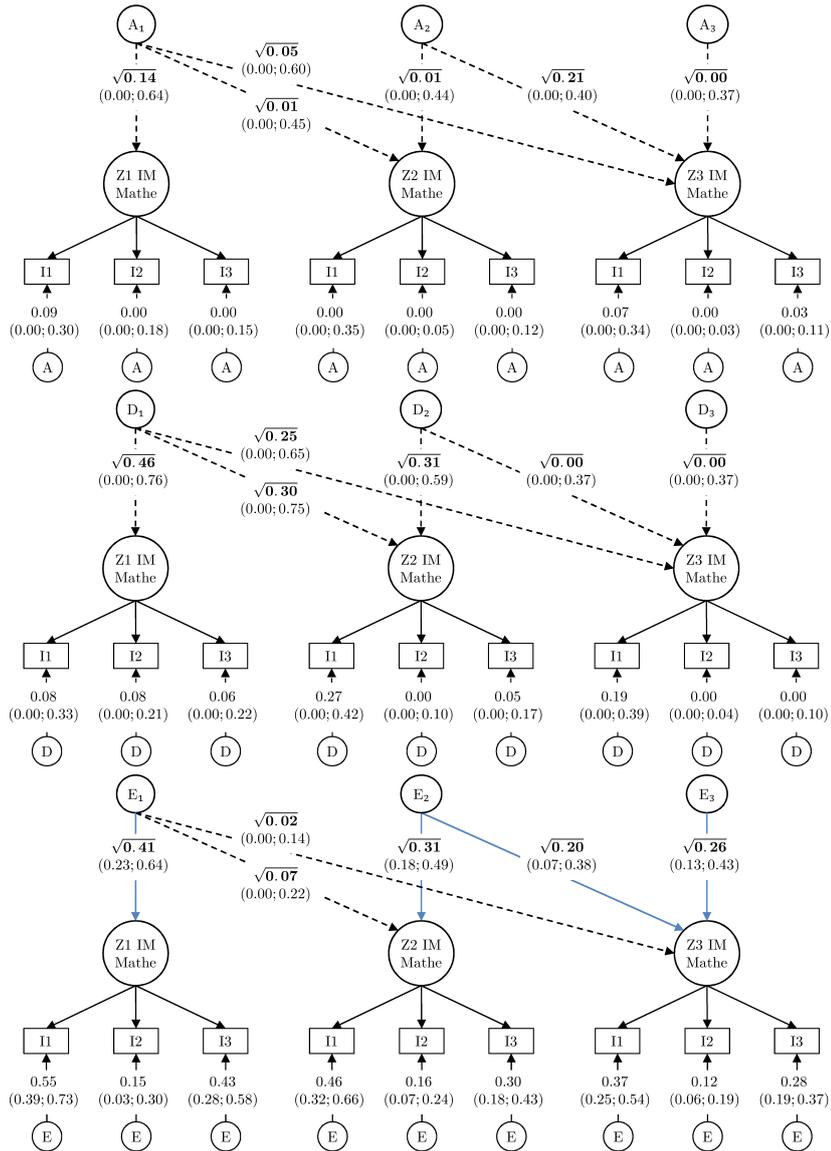


Abbildung 14.4. Cholesky Zerlegung des intrinsischen Motivation Mathe-
 matik. Die Pfadschätzungen sind aus Gründen einer einfacheren Interpretierbarkeit als Wurzel des standardisierten partiellen Regressionsgewichtes dargestellt. Die Angabe in Klammern unter der Pfadschätzung gibt das 5% Konfidenzintervall der Schätzung an.

FSK und IM Deutsch

Für die verbleibenden Variablen FSK und IM Deutsch wurden ACE Modelle geschätzt. Wie aus Tabelle 14.6 hervorgeht, war der Fit für beide Modelle mit einem RMSEA von 0.13 sehr schlecht. Die resultierenden Pfadschätzungen sind daher nur als Tendenz interpretierbar (siehe Abbildung 14.5). Insbesondere für das FSK lag eine starke genetische Komponente zu Z1 vor, die zudem nahezu die gesamte genetische Varianz zu Z2 und Z3 erklärte. Auch bei der IM Deutsch (siehe Abbildung 14.6) zeigte sich dieses Muster, allerdings waren hier die genetischen Effekte insgesamt niedriger. Im Gegensatz dazu

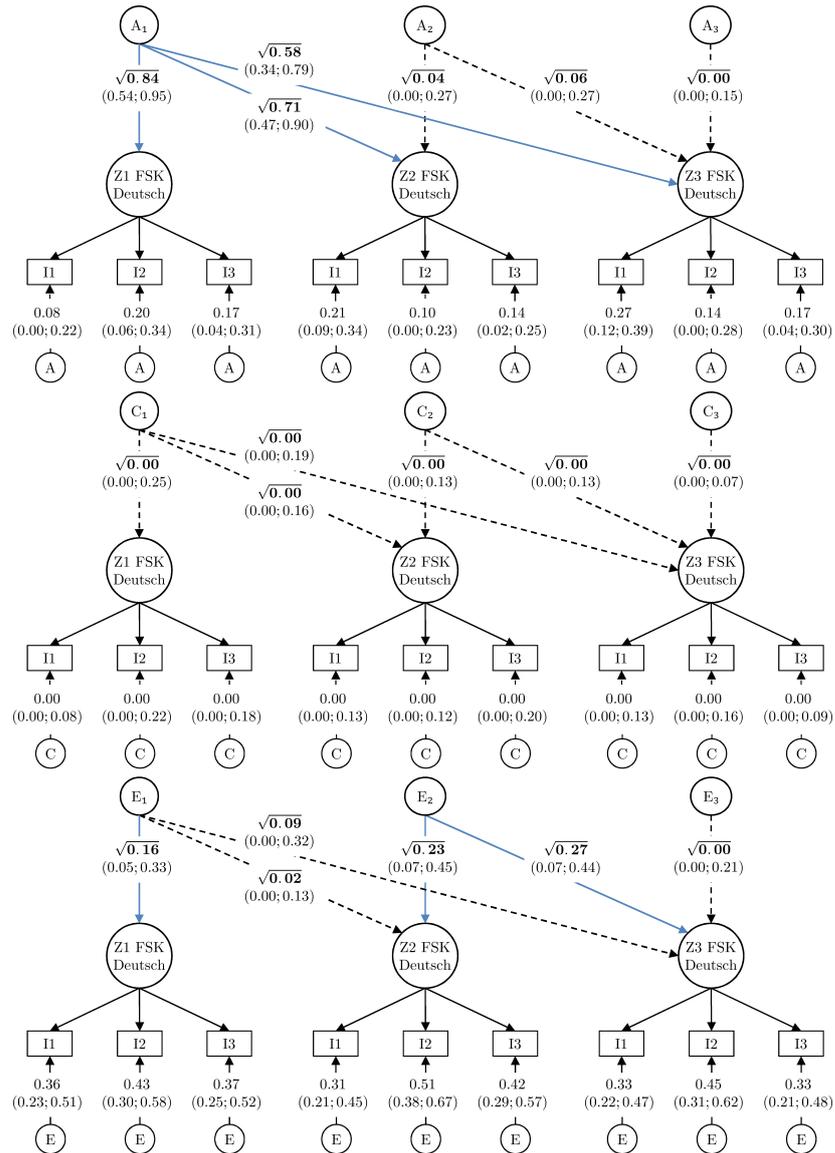


Abbildung 14.5. Cholesky Zerlegung des Fähigkeitsselbstkonzeptes Deutsch. Die Pfadschätzungen sind aus Gründen einer einfacheren Interpretierbarkeit als Wurzel des standardisierten partiellen Regressionsgewichtes dargestellt. Die Angabe in Klammern unter der Pfadschätzung gibt das 5% Konfidenzintervall der Schätzung an.

waren die geteilten Umwelteffekte zu allen Zeitpunkten und für beide Motivationskonstrukte eher unbedeutend mit Ausnahme der IM Deutsch zu Z_1 ($c^2 = 0.13$). Bei der nichtgeteilten Umwelt zeigte sich ein zu Mathematik ähnliches Bild. Die nichtgeteilten Umwelteffekte waren vor allem für IM messzeitpunktspezifisch. Sowohl im FSK als auch der IM lagen zwischen Z_2 und Z_3 überlappende nichtgeteilte Umwelteffekte vor (75% bzw. ca. 30%). Auch bei der bivariaten Erblichkeit ergab sich sowohl für das FSK als auch für die IM eine größtenteils genetische Vermittlung der phänotypischen Korrelation (zwischen 64% und 93%), die verbleibende phänotypische Korrelation wurde durch die nichtgeteilte Umwelt vermittelt.

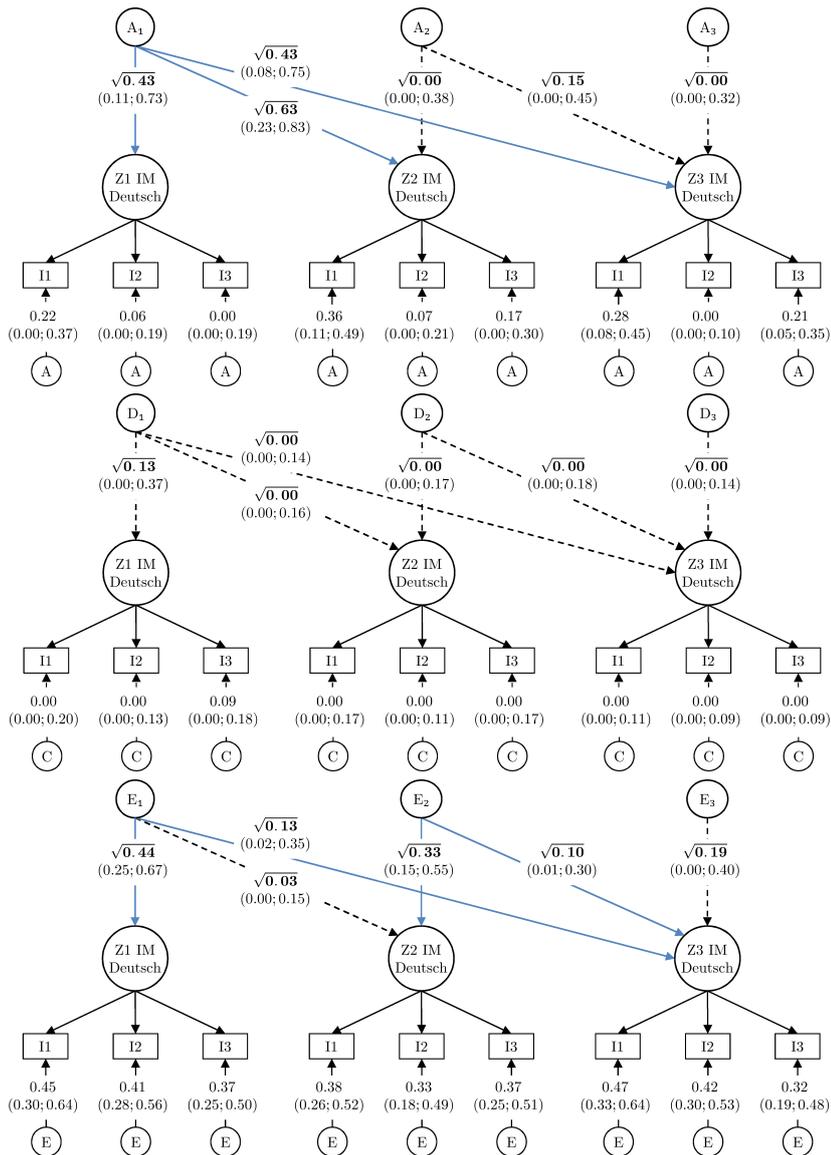


Abbildung 14.6. Cholesky Zerlegung der intrinsischen Motivation Deutsch. Die Pfadschätzungen sind aus Gründen einer einfacheren Interpretierbarkeit als Wurzel des standardisierten partiellen Regressionsgewichtes dargestellt. Die Angabe in Klammern unter der Pfadschätzung gibt das 5% Konfidenzintervall der Schätzung an.

14.2.2 Verhaltensgenetische Cross-Lagged Analyse

Als abschließende ätiologische Analyse wurden die aus der phänotypischen Betrachtung hervorgegangenen potenziellen Entwicklungstrends in der Vorhersage schulischen Erfolgs durch FSK und IM untersucht. Wie aus den Darstellungen in Kapitel 13 hervorgeht, war ein solcher Effekt einzig für den SE und das FSK in Mathematik festzustellen. Aus diesem Grund beschränkte sich die verhaltensgenetische CL-Analyse auf den Pfad SE in der 4.Klasse auf FSK in der 6.Klasse. Die hierfür verwendete Stichprobe war die

Gleiche wie in der phänotypischen Analyse (zusammengefasste Daten aus K1 und K3). Den Analysen wurden dementsprechend 65 EZ Paare und 104 DZ zugrunde gelegt.

Das Vorgehen bei der Analyse dieses Pfades erfolgte wie unter Kapitel 10 beschrieben. Für den interessierenden CL-Pfad wurde dementsprechend der relative Anteil der genetischen bzw. geteilten und nichtgeteilten Umweltvarianz an der phänotypischen Korrelation bestimmt. Als vollständiges Modell wurde ein ACE Modell gewählt, obgleich SE und FSK eine unterschiedliche Ätiologie (ACE respektive ADE) aufwiesen. Dieses Vorgehen wurde gewählt, weil davon ausgegangen werden kann, dass bei einer Nichtberücksichtigung von eigentlich vorliegenden nicht-additiven genetischen Effekten diese den additiven genetischen Effekten zugerechnet werden. Eine Nichtberücksichtigung von geteilten Umwelteffekten würde allerdings eine Überschätzung genetischer Effekte zur Folge haben (Y. L. L. Luo et al., 2011).

Die phänotypische Korrelation ergab sich zu:

$$a_{22}a_{42} + c_{22}c_{42} + e_{22}e_{42} = \sqrt{0.44 \cdot 0.04} + \sqrt{0.00 \cdot 0.01} + \sqrt{0.03 \cdot 0.10} = 0.20$$

Diese wurde zum überwiegenden Teil durch additive genetische Effekte vermittelt (70%; $(a_{22}a_{42})/0.20$), der verbleibende Anteil ging nahezu vollständig auf nichtgeteilte Umwelteffekte zurück (29%; $(e_{22}e_{42})/0.20$). Der Anteil der geteilten Umwelt lag lediglich bei 1% ($(c_{22}c_{42})/0.20$).

Nach dieser Darstellung der Ergebnisse der durchgeführten phänotypischen sowie ätiologischen Ergebnissen zur Beantwortung der Fragestellungen dieser Arbeit, erfolgt im anschließenden Diskussionsteil eine zusammenfassende Interpretation und Diskussion der Befunde.

VI. DISKUSSION

Im Rahmen der vorliegenden Arbeit wurde die zentrale Fragestellung der Vorhersage schulischen Erfolgs durch motivationale Prädiktoren unter Kontrolle von allgemeinen kognitiven Fähigkeiten sowohl phänotypisch als auch ätiologisch untersucht. Die phänotypischen Analysen gliederten sich dabei in zwei große Bereiche: Zunächst wurde die intra- und interindividuelle Stabilität der Prädiktoren über den Untersuchungsverlauf hinweg betrachtet. Hiernach wurde die Vorhersage schulischen Erfolgs in den Fächern Mathematik und Deutsch in einem längsschnittlichen MKSD modelliert, wobei auch das Vorliegen von potenziellen Entwicklungs- und Kohorteneffekten geprüft wurde. Die Kontrolle allgemeiner kognitiver Fähigkeiten diente schließlich der Abschätzung der inkrementellen Validität der motivationalen Variablen.

In Erweiterung der phänotypischen Betrachtung wurden zudem in univariaten und multivariaten genetischen Analysen die Ursachen interindividueller Unterschiede in den Variablen sowie von Stabilität und Veränderung untersucht. Schließlich wurde der gefundene Entwicklungseffekt für das FSK in Mathematik hinsichtlich einer genetischen und umweltbedingten Beeinflussung geprüft.

Die nachfolgende Diskussion der unter Kapitel V dargestellten Ergebnisse erfolgt analog zu dieser Einteilung in phänotypische und ätiologische Befunde und orientiert sich an den in Kapitel III formulierten Hypothesen.

15 Diskussion der Ergebnisse aus der phänotypischen Betrachtung

15.1 Intra- und interindividuelle Stabilität

In der phänotypischen Betrachtung interessierte zunächst die intra- und interindividuelle Stabilität der Variablen über den Untersuchungsverlauf hinweg. Es wurde davon ausgegangen, dass die latenten Mittelwerte des schulischen Erfolgs sowie der motivationalen Variablen über die Zeit hinweg abnehmen, bei gleichzeitiger Zunahme an Varianz. Außerdem wurde eine hohe interindividuelle Stabilität postuliert. Die nachfolgende Darstellung fasst die diesbezüglichen Ergebnisse zusammen und interpretiert deren Bedeutung im Kontext des empirischen Forschungsstandes. Für den latenten Faktor der allgemeinen kognitiven Fähigkeiten konnten diese Annahmen aufgrund der fehlgeschlagenen Modellierung nicht geprüft werden. Mögliche Gründe hierfür werden nachfolgend kritisch diskutiert.

15.1.1 Schulerfolg in Mathematik und Deutsch

In Bezug auf eine Veränderung der schulischen Leistung in Mathematik und Deutsch von der Grundschule in die weiterführende Schule ließen die Ergebnisse der Analyse der latenten Mittelwerte in der Modellschätzung nicht auf einen negativen Verlauf schließen. Allerdings fand sich bei Gleichsetzung der latenten Varianzen über die Zeit eine signifikante Verringerung des Modell-Fits, v.a. im Schulfach Mathematik, die Varianz der Schulnoten konnte also nicht als gleich angenommen werden. Hypothese 1a gilt dementsprechend für den SE nur in Bezug auf die angenommene steigende Varianz im Schulfach Mathematik als bestätigt.

Unabhängig von der Modellschätzung können die latenten Mittelwerte auch numerisch betrachtet werden. Dabei fällt zunächst auf, dass diese insgesamt (in allen Kohorten und zu allen Messzeitpunkten) in einem hohen Bereich lagen. Die Kinder der untersuchten Stichprobe zeigten also sowohl in Mathematik als auch in Deutsch im Mittel gute bis sehr gute Leistungen. Dabei waren die Noten in der Grundschule erwartungsgemäß

höher als in der weiterführenden Schule und fielen für die jüngeren Kinder in der Regel auch etwas besser aus. Es ließ sich außerdem beobachten, dass die Noten in Deutsch etwas schlechter waren als die Noten in Mathematik. Weiterhin konnte gezeigt werden, dass die latenten Mittelwerte im Untersuchungsverlauf, und hier insbesondere nach dem Übergang auf die weiterführende Schule, etwas absanken, dieser Effekt war jedoch nicht ausgeprägt genug, um in der Modellschätzung Signifikanz zu erreichen. Auch war festzustellen, dass die Schulnoten in der Erhebung nach dem Wechsel wieder etwas anstiegen.

Ergänzend soll an dieser Stelle noch die manifeste Interkorrelation der Noten in Mathematik und Deutsch innerhalb eines Schuljahres näher betrachtet werden, da diese für die weitere Interpretation der Ergebnisse und die Einbettung der Arbeit in den Kontext vorhandener empirischer Evidenzen relevant sind.

Die Interkorrelationen wurden auf Basis einer aus Halbjahres- und Jahresendnote gemittelten Gesamtnote berechnet. Generell lag der Zusammenhang zwischen den Noten innerhalb eines Schuljahres in einem moderaten Bereich zwischen $r = 0.43$ bis $r = 0.62$. Dabei zeigte sich auch in der Höhe der Interkorrelationen eine Veränderung über den Untersuchungsverlauf hinweg, denn der Zusammenhang der Mathematik- und Deutschleistung war in den Grundschuljahren höher, als dies auf der weiterführenden Schule der Fall war. Dies bedeutet zunächst, dass diejenigen Kinder, die gute Leistungen in Mathematik erbringen, in der Regel auch in Deutsch vergleichbar gute Leistungen zeigen. Allerdings differenziert sich dieser fächerübergreifende hohe schulische Erfolg mit zunehmendem Alter. Auf der weiterführenden Schule zeigt sich eher eine Leistungs-differenzierung dahingehend, dass Kinder nun häufiger entweder in Mathematik oder Deutsch bessere Leistungen zeigen. Bei Betrachtung eines längeren zeitlichen Intervalls auf der Sekundarschule würde sich dies wahrscheinlich noch deutlicher herauskristallisieren.

Die zuvor referierten Ergebnisse der vorliegenden Untersuchung lassen sich nur in Teilen mit vorhandenen empirischen Ergebnissen in Einklang bringen. Andere Studien konnten hier deutlicher ein konstantes Absinken der schulischen Leistung in Form von Schulnoten belegen (vgl. z. B. Schneider et al., 1997), dies galt insbesondere im Schuljahr nach dem Wechsel auf die Sekundarschule und hier vor allem für Gymnasialschüler (vgl. z. B. Ball et al., 2006; Blyth et al., 1983). Aus diesem Grund liegt die Vermutung nahe, dass sich in der hier vorliegenden Stichprobe, in der ein überwiegender Teil der Kinder das Gymnasium besuchte (vgl. dazu Kapitel 7.2.2), ein besonders starkes Absinken der Mittelwerte in Mathematik und Deutsch finden würde.

Eine mögliche Ursache dieser Diskrepanz zu anderen Studien könnte in der hier gewählten latenten Modellierung schulischen Erfolgs mit korrelierten Messfehlern liegen. Durch die Berücksichtigung von Messfehlern ist die latente Messung nicht mehr direkt

vergleichbar mit einer manifesten, was sich in höheren latenten Werten manifestiert. Wird davon ausgegangen, dass der Fehler in den Noten unterschiedlicher Schuljahre Ähnliches umfasst, führt diese Kontrolle der Fehlervarianz zu ähnlicheren latenten Messwerten als in einer manifesten Skalenbildung. Hinzu kommt vor dem Hintergrund der längsschnittlichen Betrachtung ab der zweiten Klassenstufe¹, dass insbesondere in den frühen Schuljahren (Z1) mehr Daten mit einer ML-Schätzung ersetzt werden mussten als zu späteren Messzeitpunkten. Da diese Ersetzung auf Basis der vorhandenen Daten erfolgt, werden hierdurch die Werte einander ähnlicher als sie dies eventuell in der Realität wären.

Die beobachteten hohen Werte in den Schulnoten decken sich jedoch auch mit den Befunden anderer Studien die sich auf den Grundschulbereich beziehen (vgl. z. B. Tent, 2006; H. S. Weber, 2012) und auch die hohe Interkorrelation der Schulnoten in Mathematik und Deutsch fügt sich in das Bild eines multipel begabten (Grund-)Schülers ein. Kritisch angemerkt sei an dieser Stelle die bereits angesprochene Selektivität der KoSMoS Stichprobe hinsichtlich des SES. Zahlreiche Studien konnten in diesem Zusammenhang positive Korrelationen zwischen SES und Intelligenz aufzeigen (vgl. z. B. Hanscombe et al., 2012). In Anbetracht der hohen Erblichkeit von Intelligenz (vgl. z. B. T. Bouchard & McGue, 1981; Deary et al., 2006; Segal & Johnson, 2009), kann denn auch bei Kindern aus Familien mit hohem SES in der Regel von einer höheren Intelligenz ausgegangen werden, die dann wiederum in positiver Relation zum schulischen Erfolg von Kindern steht (Gustafsson & Undheim, 1996). Hinzu kommt, dass verschiedene Studien auch in Bezug auf die Bildungsaspiration und die Häufigkeit der Gymnasialempfehlung positive Korrelationen mit dem SES nachgewiesen haben (Ditton, 2010b). Dies bedeutet, dass Kinder aus Arbeiterfamilien mehr Leistung erbringen müssen als Kinder aus Familien, deren Eltern einen akademischen Hintergrund aufweisen. Nichtsdestotrotz lagen auch in Studien mit hinsichtlich des SES repräsentativeren Stichproben hohe Korrelationen zwischen den Leistungen in unterschiedlichen Fächern in der Grundschule vor (H. S. Weber, 2012). Im Kontext der vorliegenden Untersuchung wurde jedoch deutlich, dass sich dies auf der Sekundarschule in Richtung eines differenzierteren Leistungsprofils verschiebt. Hier wäre es interessant, auch die Zusammenhänge zu weiteren Schulfächern zu untersuchen, die einmal eher im Spektrum mathematisch-naturwissenschaftlicher Kompetenzen und einmal eher in sprachlichen Kompetenzen angesiedelt sind.

Ein genauere Betrachtung der Varianz der Schulnoten liefert ebenfalls ein differenzierteres Gesamtbild als die bloße Bewertung des Modell-Fits. Zunächst kann festgehalten werden, dass die Varianz der Noten insgesamt eingeschränkt

¹ In vielen Schulen werden erst ab Ende der zweiten Klassenstufe Zensuren vergeben, weshalb für viele Kinder von K1 keine Halbjahresnote zu Z1 vorlag.

war, obwohl eine Gleichsetzung der Varianzen im Modellvergleich nicht möglich war. Die latente Varianz in den Mathematik- und Deutschnoten lag in allen Kohorten und in allen Schuljahren in einem Bereich von weniger als einer Notenstufe. Auch in Bezug auf die Varianzen fanden sich schulfachspezifische Effekte, die allerdings aufgrund der insgesamt niedrigen Varianz nur als Tendenzen interpretierbar sind und nicht überbewertet werden sollten. Für Mathematik zeigte sich, dass in den beiden älteren Kohorten zunächst eine Verringerung der Varianz von $Z1$ nach $Z2$ in den Schulnoten vorlag. Die Kinder unterschieden sich also im ersten erfassten Schuljahr nach dem Schulwechsel weniger stark in ihrer Mathematikleistung als in der Grundschule. Erst im weiteren Verlauf der Schullaufbahn stieg die latente Varianz nochmals an und lag dann in einem höheren Bereich als zu $Z1$. Für die Kinder der ersten Kohorte erhöhte sich hingegen die Varianz direkt nach dem Übergang (von 0.37 nach 0.75), wobei hier einschränkend zu beachten ist, dass in dieser Kohorte zu $Z1$ meist nur Jahresendnoten vorlagen und aufgrund dessen in K1 besonders viele Noten ersetzt werden mussten. Im Schulfach Deutsch zeigte sich entgegen der Erwartung eine Verringerung der Varianz über die Schuljahre hinweg. Dieser Effekt wird im Folgenden mit Blick auf die noch geringere Variabilität v.a. zu $Z2$ und $Z3$ (ca. 0.35) inhaltlich nicht weiter interpretiert. Dieser Befund einer höheren Veränderung in den Mathematik- im Vergleich zu den Deutschnoten wurde bereits in anderen Untersuchungen gezeigt (vgl. z. B. H. S. Weber, 2012). In Mathematik treten also eher Leistungsveränderungen auf als in Deutsch, was eventuell damit in Zusammenhang steht, dass in Mathematik unterschiedlichere Anforderungen an die Kinder gestellt werden als dies in Deutsch der Fall ist. In Deutsch fließen in allen Schuljahren ähnliche Kompetenzen in die Notengebung ein (beispielsweise Rechtschreibung, Lesekompetenz), während sich die Note in Mathematik gezielter auf eine bestimmte Fähigkeit bezieht (beispielsweise Dividieren, Bruchrechnen).

Es stellt sich nunmehr die Frage, warum sich in Mathematik zunächst eine Verringerung der Varianz nach dem Wechsel in die weiterführende Schule fand, sich die Kinder also in ihren Leistungen ähnlicher wurden, und erst danach eine breitere Fächerung der Noten zu finden war. Die ähnlicheren Leistungen der Kinder in den beiden ersten Jahren nach dem Wechsel auf die Sekundarschule könnten eventuell mit den Unterrichtsinhalten in Mathematik² in Zusammenhang stehen. So könnte vermutet werden, dass auf dem Gymnasium zunächst alle Kinder auf ein gemeinsames Niveau in den mathematischen Kompetenzen gebracht werden sollen. Hierdurch könnte gewährleistet werden, dass die für komplexere mathematische Operationen notwendigen Grundlagen bei allen Kindern vorhanden sind. Wenn diese Grundlagen nun bereits in der Grundschule Unterrichtsgegenstand wären, könnte diese Wiederholung des

² In der Folge wird sich auf den Lehrplan der Gymnasien bezogen, weil ein Großteil der Kinder diese Schulform besuchte.

Stoffes in den ersten Gymnasialschuljahren zu einer Angleichung der mathematischen Leistung der Kinder führen.

Näheren Aufschluss liefert hier der Lehrplan in Mathematik des Landes Nordrhein-Westfalen³ für die Grundschule und die ersten beiden Schuljahre des Gymnasiums. Hieraus geht hervor, dass sich in den für das Gymnasium definierten Kernkompetenzbereichen Algebra/Arithmetik, Funktionen, Geometrie und Stochastik viele, bereits in der Grundschule als Kompetenzerwartung am Ende der 4. Klasse umschriebene Fähigkeiten wiederfinden.^{4,5} Es könnte entsprechend in der Tat einen solchen Wiederholungseffekt geben, der sich erst im weiteren Schulverlauf wieder verliert. Werden in den höheren Stufen dann höhere mathematische Operationen Unterrichtsinhalt, könnte dies wieder in einer größeren Varianz der mathematischen Schulleistung resultieren. Hier ist auch denkbar, dass sich bei einer Stichprobe mit dispergierenderen Fähigkeiten andere Muster finden ließen.

Eine weitere Erklärung für die höhere Variabilität der Mathematiknote könnte auch darin zu finden sein, dass die Noten in beiden Fächern unterschiedlich objektiv sind. Während in Mathematik eine richtige und falsche Lösung zu identifizieren ist, können in Deutsch auch Verhaltensaspekte (z. B. Beteiligung am Unterricht, Fleiß, Kreativität oder Aufmerksamkeit) in die Notengebung einfließen, auch wenn Dinge wie grammatikalische Richtigkeit und Rechtschreibung ebenfalls eindeutig zu klassifizieren sind. Dennoch ergeben sich hier für den Lehrer mehr Möglichkeiten, auch Aspekte wie Ideenreichtum o.ä. zu berücksichtigen. Folglich wäre es in Deutsch leichter, durch Motivation und Anstrengungsbereitschaft gewisse kognitive Defizite auszugleichen.

In einem weiteren Schritt wurde die interindividuelle Stabilität in den Schulnoten betrachtet. Hier zeigte sich erwartungskonform eine außerordentlich hohe Rangreihenstabilität sowohl für die jüngeren als auch die älteren Kinder. Diese fiel bei Berücksichtigung des Messfehlers in der latenten Modellierung wesentlich höher aus als in der manifesten Betrachtung. Hypothese 1b kann also für den schulischen Erfolg in Mathematik und Deutsch als bestätigt gelten. Im Speziellen zeigte sich stets eine höhere Stabilität zwischen zeitlich benachbarten Erhebungszeitpunkten als zwischen Z1 und Z3. Aber selbst in diesem längsten Zeitintervall von insgesamt sechs Jahren fanden sich beachtliche Stabilitäten von bis zu $r_{tt} = 0.69$. Bei differenzierter Betrachtung der Stabilität der Noten wurde weiterhin deutlich, dass in der Folge des Wechsels auf die weiterführende Schule (in K2 und K3 zwischen Z1 und Z2) zunächst eine geringere Sta-

³ Ein Großteil der KoSMoS Stichprobe wurde in NRW rekrutiert.

⁴ <http://www.standardsicherung.schulministerium.nrw.de/lehrplaene/kernlehrplaene-seki/gymnasium-g8/>

⁵ <http://www.standardsicherung.schulministerium.nrw.de/lehrplaene/lehrplaene-gs/mathematik/lehrplan-mathematik/kompetenzen/>

bilität der Noten vorlag als im Verlauf der Grundschule (zu beobachten in K1 von Z1 nach Z2) bzw. der weiterführenden Schule (zu beobachten in K2 und K3 zwischen Z2 und Z3). Der Wechsel in eine andere Schulform geht somit, wie nicht anders zu erwarten, zunächst mit einer gewissen Verschiebung der Rangreihe in schulischen Leistungen einher, im weiteren Verlauf bleibt die Rangreihe dann in großen Teilen unverändert bestehen.

Diese Ergebnisse stehen in Einklang mit Befunden aus der Literatur, die übereinstimmend eine hohe Stabilität von bis zu 0.80 in den Schulnoten belegen (Böhnel, 1996; Tent, 2006). Die gefundene geringere Stabilität nach dem Wechsel auf die weiterführende Schule könnte auf die von den Kindern mit dem Schulübergang geforderte soziale sowie leistungsbezogene Anpassung zurückzuführen sein (Büchner & Koch, 2001; Koch, 2004). In diesem Zusammenhang wäre eine Analyse möglicher Moderatoren einer gelungenen Anpassung, beispielsweise des Involvements bzw. des Erziehungsstils der Eltern oder der Copingressourcen des Kindes (vgl. z. B. Crockett et al., 1989; Duchesne et al., 2005), auch im Hinblick auf mögliche Interventionen sicherlich lohnenswert.

15.1.2 Allgemeine kognitive Fähigkeiten

Wie bereits aus der Ergebnisdarstellung hervorging, konnten die Fragestellungen hinsichtlich der intra- und interindividuellen Stabilität allgemeiner kognitiver Fähigkeiten aufgrund der fehlgeschlagenen latenten Modellierung des ersten g -Faktors im Rahmen der vorliegenden Arbeit nicht beantwortet werden. An dieser Stelle erfolgt daher eine kritische Diskussion der Operationalisierung kognitiver Fähigkeiten, insbesondere des ersten Messzeitpunktes.

Die Operationalisierung der kognitiven Fähigkeiten wurde in KoSMoS analog zur englischen Zwillingsstudie TEDS gestaltet und bestand aus zwei verbalen und zwei non-verbalen Tests (siehe Kapitel 8.3). Zum ersten Messzeitpunkt im Jahr 2005 wurde auf eine elternadministrierte Heimtestung zurückgegriffen, zur zweiten Erfassung der kognitiven Fähigkeiten fand eine Umstellung auf eine telefonbasierte Testung durch geschulte Testleiter statt (vgl. Kapitel 8.3.1). Des Weiteren erhielten alle Kinder in Z1 die gleichen Items, während zu Z3 altersspezifische Items vorgegeben wurden. Beide Vorgehensweisen (Heimtestung und Telefontestung) hatten sich dabei in der englischen Stichprobe als adäquate Methoden zur Testung kognitiver Fähigkeiten bei geografisch weit verstreut liegenden Stichproben, bei denen keine Face-to-Face Testung möglich ist, bewährt (Petrill et al., 2002; Toussaint, 2007). Die Entscheidung zu einem Wechsel der Erhebungsmethode innerhalb von KoSMoS basierte letztlich auf mehreren empi-

rischen Befunden.⁶ Zum einen blieb die Korrelation eines manifesten g -Faktors mit den Schulnoten zu $Z1$ mit 0.33 für Mathematik und 0.32 für Deutsch deutlich hinter den aus der Literatur zu erwartenden Korrelationen von ca. 0.50 zurück (Toussaint, 2007). Auch die bei Toussaint (2007) und Spengler (2009) berichteten univariaten verhaltensgenetischen Analysen des g -Faktors zu $Z1$ deuteten auf einen von der Literatur etwas abweichenden genetischen Einfluss hin, wenngleich diese Abweichung für die betrachtete Altersgruppe von Kindern im Grundschulalter in einem durchaus plausiblen Rahmen lagen.

Weitere Belege für die Sinnhaftigkeit einer Umstellung der Erhebungsmethode lieferten die bereits unter Kapitel 8.3 vorgestellten Arbeiten von Pooch (2006) sowie Marthaler (2007). Diese sprachen zum einen für eine verminderte Objektivität der Heimtestung aufgrund des Hilfeverhaltens der Eltern und einer daraus resultierenden geringeren Validität. So ergab sich bei Pooch (2006) eine nur moderate Korrelation ($r = 0.58$) zwischen elternadministrierter Heimtestung und einer Face-to-Face Testung und auch der Zusammenhang zu Schulnoten lag mit $r = 0.33$ in einem niedrigeren Bereich als die Face-to-Face Testung ($r = 0.66$). Hingegen konnte Marthaler (2007) zeigen, dass eine Telefontestung wesentlich höher mit einer Face-to-Face Testung korrelierte ($r = 0.81$).

Durch die Aufschlüsselung der gesamten KoSMoS Stichprobe in drei Alterskohorten, welche in der vorliegenden Studie vorgenommen wurde, konnte dieses Bild für die kognitive Messung zu $Z1$ weiter differenziert werden. Wie in der Gesamtstichprobe konnte auch hier auf Basis einer durchgeführten explorativen Faktorenanalyse ein starker erster unrotierter g -Faktor in allen Kohorten generiert werden. Die vier Subtests luden auf diesem Faktor substantiell und in ähnlicher Höhe. Auch die Reliabilität der Subtests lag nach Korrektur bis auf wenige Ausnahmen in einem guten bis sehr guten Bereich ($r_{tt_corr} = 0.47$ bis $r_{tt_corr} = 0.95$). In der latenten Modellierung wurde hingegen deutlich, dass der Fit insbesondere in K2 nicht ausreichend hoch war. Die Ladungen der vier Subtests auf einen latenten g -Faktor lagen in einigen Fällen nur knapp über 0.30, was als Hinweis darauf gewertet werden kann, dass die Subtests weniger Varianz teilen, als ursprünglich intendiert. Aus diesem Grund fiel auch die Fehlervarianz der Subtest hoch aus. Dieser Effekt zeigte sich im Übrigen auch in der latenten Schätzung des g -Faktors zu $Z3$. Hier wurde deutlich, dass vor allem die beiden verbalen Untertests niedrigere Ladungen auf den gemeinsamen latenten Faktor aufwiesen.

Diese kohortenspezifisch differierende Passung des latenten Faktors setzte sich auch im Zusammenhang eines manifesten g -Faktors mit schulischem Erfolg fort. Auch in diesem Fall zeigte sich, dass die Korrelation mit den Schulnoten insbesondere in K2

⁶ Im Unterschied zur vorliegenden Studie wurden die nachfolgend berichteten Ergebnisse auf Basis einer Gesamtstichprobe und nicht kohortenspezifisch gewonnen.

sehr niedrig ausfiel, während sie in K1 und K3 zumindest in einem moderaten Bereich ($r = 0.25$ bis $r = 0.37$) lag. Dementsprechend geht die z. B. bei Toussaint (2007) berichtete, verminderte externe Validität der kognitiven Testung vor allem auf die Daten aus K2 zurück. Für den latenten g -Faktor zu Z3 lag die Korrelation erwartungskonform meist bei ca. 0.50. Hier war festzustellen, dass der Zusammenhang zur Mathematiknote etwas höher ausfiel als derjenige zur Deutschnote, was im Einklang mit Ergebnissen anderer Studien steht (Amthauer et al., 2001). Erstaunlicherweise lag insbesondere für das Schulfach Mathematik die Korrelation zwischen dem g -Faktor zu Z3 und der Schulnote zu Z1 in K1 und K3 höher als mit den zeitgleichen Noten zu Z3. Zudem war auch zu konstatieren, dass die Beziehung zwischen den kognitiven Fähigkeiten und dem FSK nur in K3 den aus der Literatur zu erwartenden Zusammenhängen entsprachen (z. B. B. Spinath et al., 2006). Auch dieser Aspekt kann als mangelnde externe Validität des erhobenen g -Maßes gewertet werden. Eine aktuelle Studie erweitert die kognitive Testbatterie um zwei quantitative Fähigkeitstests des KFT 4–12+R (Heller & Perleth, 2000). In dieser Studie mit Schülern der fünften Klasse erwies sich der Modell-Fit eines latent modellierter g -Faktors als ausgezeichnet (M. J. Frey, 2012). Obgleich diese Testung als Face-to-face Testung durchgeführt wurde, liefert sie dennoch einen Hinweis darauf, dass eventuell die vier eingesetzten Untertests, und hier insbesondere die verbalen Verfahren, die auch bei M. J. Frey (2012) die geringsten Ladungen auf dem g -Faktor aufwiesen, als Indikator kognitiver Fähigkeiten nicht ausreichend waren. Hier sollte in Folgeuntersuchungen die Möglichkeit einer Erweiterung der kognitiven Testbatterie geprüft werden.

Vor diesem Hintergrund scheint natürlich auch die Frage auf, warum sich die kognitive Erhebung gerade in K2 als inadäquat erwiesen hat. Um dies näher zu beleuchten, wurden mögliche Erklärungen einer detaillierteren Betrachtung unterzogen. Es sei an dieser Stelle zunächst nochmal erwähnt, dass die zu Z1 vorgegebenen Items aus dem KFT 4-12+R diejenigen Items umfasste, die für 4. Klassen vorgesehen waren und diese demzufolge für K3 am besten passen sollten. Für K1 sollten diese Items folglich zu schwer gewesen sein. Zu Z3 erfolgte die Itemvorgabe dann altersentsprechend.

Zunächst wurden hierzu die Kohorten hinsichtlich ihrer deskriptiven Kennwerte vergleichend betrachtet. Hier zeigte sich, dass die Varianz in K2 zwar stets am niedrigsten ausfiel, die Mittelwerte und Standardabweichungen zu beiden Messzeitpunkten aber insgesamt vergleichbar waren mit denen von K1 und K3. Auch in Bezug auf die Reliabilität und der Anzahl an nicht gelösten Items konnten keine Unterschiede zwischen den Kohorten festgestellt werden. Ein Ausschluss von Ausreißern führte ebenfalls zu keiner Veränderung der Ergebnisse. Ein Unterschied zeigte sich hingegen in der Interkorrelation der nonverbalen Skalen, die in Z1 für K2 niedriger lagen als für K1 und K3. Zu Z3 war dies nicht mehr der Fall. Auch in der Stabilität der Einzeltests ergab sich in K2 eine

Inkonsistenz im Vergleich zu K1 und K3: Hier lag die errechnete Stabilität des Subtests *Figurenklassifikation* mit $r_{tt} = 0.05$ nahe Null. Dies bedeutet, dass diejenigen Kinder, die zu Z1 gut abschnitten zu Z3 schlechtere Leistungen zeigten und umgekehrt, wobei dies in dieser extremen Ausprägung ausschließlich in K2 vorkam. Hieran knüpfte die Vermutung an, dass sich eventuell gerade für diese Kinder ein deutlicherer Unterschied in den Items der ersten Messung im Vergleich zur altersspezifischen Itemvorgabe zu Z3 ergeben haben könnte, die Items sich also unähnlicher waren als dies für K1 und K3 der Fall gewesen war. Um diesen Punkt zu prüfen wurden Skalen berechnet, die auch für Z3 nur diejenigen Items enthielten, die in allen Kohorten vorgegeben wurden.⁷ Es zeigten sich jedoch auch in diesen Skalen niedrigere Stabilitäten für K2, weshalb nicht davon ausgegangen werden konnte, dass die Items für K2 sich diesbezüglich maßgeblich von den Items von K1 und K3 unterschieden.

Wenngleich dies nicht in unmittelbarem Zusammenhang mit der divergenten Verteilung der kognitiven Fähigkeiten in K2 zu sehen war,⁸ wurden die beobachteten Deckeneffekte nochmals einer genaueren Betrachtung unterzogen. Ein solcher Deckeneffekt konnte in erster Linie zu Z1 festgestellt werden, es wurden also viele Items von allen Kinder korrekt gelöst (vgl. dazu Kapitel 11.2). Hier zeigte sich denn auch, dass die Mittelwerte der beiden nonverbalen Untertests in allen Kohorten zu Z3 etwas niedriger ausfielen als zu Z1, zu Z3 war die Zahl an richtig gelösten Items also niedriger als zu Z1. Bei den verbalen Subtests stiegen die Werte von K1 in beiden Skalen an, während die Werte der beiden älteren Kohorten sanken. Einen Hinweis darauf, dass die vorgegebenen Tests zu Z1 wahrscheinlich für K2 und K3 zu einfach waren, fand sich auch bei Betrachtung der Verteilungskennwerte der Skalen. Zu Z1 lag der Mittelwert in K1 in allen Untertests immer am niedrigsten, die Werte von Schiefe und Exzess zeigten in dieser Kohorte jedoch am ehesten das Vorliegen einer Normalverteilung an. Die Items differenzierten also in der jüngsten Kohorte noch am besten. Eine Inspektion der Scatterplots ergab schließlich vor allem in den nonverbalen Skalen Deckeneffekte. Dass die Items in Z3 in ihrer Schwierigkeit eher angemessen waren, zeigte sich schließlich darin, dass nun K3 in beiden nonverbalen Tests schlechter abschnitt als K1 und K2, wohingegen im Untertest *Allgemeinwissen*⁹ die Kinder aus K1, für welche die Items am schwierigsten sein sollten, auch am schlechtesten abschnitten. Das Vorliegen von Deckeneffekten ist eventuell auch auf die bereits erläuterte Selektivität der Stichprobe, und hier im Speziellen auf die Selektivität bei weiterer Teilnahme an KoSMoS (vgl. Kapitel 7.2.1), zurückzuführen.

⁷ Die Testvorgabe des KFT sieht jeweils andere Start- und Stoppitems für die unterschiedlichen Altersklassen vor, der überwiegende Teil der Items ist jedoch für die Kinder überlappend.

⁸ Hier erscheint es wenig plausibel anzunehmen, dass sich ein Deckeneffekt gerade auf die mittlere Alterskohorte und nicht auf die jüngsten bzw. ältesten Kinder in dieser Art und Weise auswirken würde.

⁹ Alle Kinder bekamen hier die gleichen Items vorgegeben.

Insgesamt betrachtet resultierte aus dieser genauen Prüfung der Verteilung der kognitiven Skalen in den Kohorten und vergleichend zwischen beiden Erhebungen, dass die vorhandene Erfassung mit einer gewissen methodischen Insuffizienz, insbesondere zu Z1, in Teilen jedoch auch zu Z3, behaftet war. Dies führte in der Folge zu einer geringeren Gewichtung der allgemeinen kognitiven Fähigkeiten im weiteren Verlauf der Studie.

15.1.3 Fähigkeitsselbstkonzept und Intrinsische Motivation

In den beiden Motivationsmaßen konnte bezüglich der Veränderung auf Mittelwertebene wie erwartet ein Absinken über die Zeit hinweg beobachtet werden, lediglich in K1 im FSK Deutsch blieben die Mittelwerte in etwa konstant. Allerdings war dieses Absinken nur in der IM so stark ausgeprägt, dass sich im Rahmen der Modellschätzung eine signifikante Veränderung ergab. Die Varianz in beiden Maßen lag, entgegen der Erwartung, insgesamt in einem niedrigen Bereich. Es war zwar eine leichte Erhöhung der Variabilität von Z1 nach Z3 in beiden Maßen zu verzeichnen, diese war jedoch, mit Ausnahme des FSKs in Mathematik für K2,¹⁰ nicht deutlich genug, um bei Gleichsetzung der Varianz zu einem schlechteren Fit zu führen.

Obwohl die in Hypothese 1a formulierten Annahmen damit nur in Bezug auf die Veränderung der Mittelwerte in der IM bestätigt werden konnten, kann ein detaillierter Blick auf die Ergebnisse und deren Relation zu vorhandenen empirischen Evidenzen aufschlussreich sein.

Die latenten Mittelwerte der Skalen des FSKs lagen in allen Kohorten und zu jedem Zeitpunkt mit Werten zwischen 3.64 und 4.32 in einem hohen Bereich. Dies bedeutet, dass die Kinder ein hoch ausgeprägtes FSK in Mathematik und Deutsch aufwiesen und die Inhalte beider Fächer auch sehr gemocht wurden. Wie erwartet lagen die Mittelwerte umso höher, je jünger die Kinder waren und sanken über den Untersuchungsverlauf numerisch betrachtet ab, d. h. die Kinder aus K1 zeigten in der Regel die höchste Ausprägung in den Skalen.

Diese hohen Mittelwerte und die in der deskriptiven Skalenstatistik deutlich gewordene Linksschiefe der Skalen weisen auf eine hohe Zustimmungstendenz hin, die gewöhnlich bei jüngeren Kindern zu beobachten ist (Möller & Trautwein, 2009; B. Spinath et al., 2006). Die hier ermittelten relativ hohen Mittelwerte im FSK korrespondieren dabei sowohl in ihrer Höhe als auch in der Standardabweichung mit denen anderer Studien (vgl. z. B. B. Spinath et al., 2006; H. S. Weber, 2012). Die gezeigte Tendenz höherer Mittelwerte vor allem bei den jüngeren Kindern der Stichprobe und das allmähliche

¹⁰ Die Varianz nahm in diesem Fall von Z1 nach Z2 jedoch zunächst ab, bevor sie nach Z3 wieder anstieg.

Absinken der Werte über die Zeit decken sich mit der Annahme, dass Kinder im frühen Grundschulalter ihre eigenen Fähigkeiten in einem sehr positiven Licht sehen (vgl. z. B. D. Stipek & Iver, 1989). Erst im weiteren Verlauf der Schulzeit und insbesondere nach dem Wechsel in die Sekundarschule, weichen diese einem realistischeren Bild der eigenen Fähigkeiten und Kompetenzen (Dweck, 2002; Guay et al., 2003; B. Spinath & F. M. Spinath, 2005a; F. M. Spinath et al., 2008a). Der negative Trend in der Höhe des FSKs wurde unter einer entwicklungspsychologischen Perspektive damit erklärt, dass Kinder mit dem Schuleintritt durch Leistungsrückmeldungen und Vergleiche mit anderen Kindern vermehrt Gelegenheit erhalten, ihre Selbstkonzepte auch durch negative Rückmeldungen zu ergänzen (Harter, 1998, 1999). Dies führt in der Folge zu realistischeren und zumeist weniger positiven Einschätzungen der eigenen Fähigkeiten (Dweck, 2002; Guay et al., 2003; Möller & Trautwein, 2009; Wigfield & Eccles, 2002). Mit Beginn des Jugendalters können Schüler ihr eigenes Stärke-Schwäche-Profil daher in der Regel recht genau einschätzen.

Es stellt sich nunmehr die Frage, warum das Absinken der Mittelwerte im FSK der Kinder in der vorliegenden Studie weniger deutlich als in anderen Studien ausgefallen ist. Als Erklärung können verschiedene methodische sowie stichprobenspezifische Aspekte angeführt werden. Zunächst greift hier bei latenter Modellierung mit korrelierten Fehlern über einen Zeitraum von sechs Jahren die gleiche Argumentation wie bereits beim schulischen Erfolg. Die Berücksichtigung der Messfehler, die in vielen Studien zur Entwicklung des FSKs nicht vorgenommen wurde, führt unter Umständen auf einer methodischen Ebene zu einer weniger starken Veränderung der Mittelwerte als bei manifester Modellierung. Die hohe Relevanz dieser Art der Modellierung wird im Übrigen auch bei Vergleich der Mittelwerte auf manifester und latenter Ebene deutlich: Zum einen waren die latenten Mittelwerte höher als die manifesten und zum anderen lagen die Mittelwerte in Mathematik auf manifester Ebene höher als diejenigen in Deutsch (ebenso wie beispielsweise bei H. S. Weber, 2012), bei latenter Modellierung war dies genau umgekehrt.

Vor dem Hintergrund des in Kapitel 3.5.1 beschriebenen BFLPEs (Marsh, 1987), wäre in der vorliegenden Untersuchung ebenfalls von einem stärkeren Absinken der Mittelwerte, speziell nach dem Übergang auf die weiterführende Schule, auszugehen gewesen. Dieser Bezugsrahmeneffekt, der in zahlreichen Studien und in verschiedenen Ländern bestätigt werden konnte (vgl. z. B. Dickhäuser & Galfe, 2004; Huguet et al., 2009; Nagegast & Marsh, 2011), würde insbesondere bei Gymnasialschülern, die mit 64% den größten Anteil der KoSMoS Stichprobe stellten, eine Verringerung des Selbstkonzeptes nach dem Übergang vorhersagen. Gerade diese Schulform sollte zahlreiche Gelegenheit zu sozialen Aufwärtsvergleichen bieten, die sich negativ auf die Einschätzung der eigenen Fähigkeit auswirken. Eine direkte Überprüfung des BFLPEs war im Rahmen des

hier vorliegenden Untersuchungsdesign jedoch nicht möglich, hierfür hätte es auch der Information über das mittlere Leistungsniveau der jeweiligen Schulklasse bedurft.¹¹ Denkbar wäre an dieser Stelle, dass bestimmte Stichprobencharakteristika dazu geführt haben, dass die Kinder der vorliegenden Stichprobe in weniger starkem Ausmaß von einer Verringerung des FSKs in Folge eines BFLPE betroffen waren. Wie zuvor bereits beschrieben lagen die Noten der Kinder in beiden Fächern auch zu Z2 und Z3 in einem hohen Bereich. Dies bedeutet, dass die Kinder wohl auch auf der weiterführenden Schule eher zu besseren Schülern zählten und die Tendenz zu einem sozialen Aufwärtsvergleich aus diesem Grund eventuell weniger stark ins Gewicht fiel. In diesem Kontext wäre es auch interessant, in ergänzenden Analysen verschiedene Schultypen innerhalb der Kohorten vergleichend zu betrachten. Im Zusammenhang mit einem stichprobenspezifischen Effekt zeigt beispielsweise Buff (1991), dass die Selbstkonzepte von Kindern in höheren Schulformen bereits in der Grundschule höher waren als solche von Kindern, die in eine andere Schulform wechselten. Auch in einer aktuellen Studie konnten Makel, Lee, Olszewki-Kubilius und Putallaz (2012) zeigen, dass in einer Gruppe von hochleistenden Schülern kein BFLPE zu beobachten war, wenn diese zusammen in einer hochleistenden Lernumwelt unterrichtet wurden. Ein vergleichbarer Effekt fand sich auch in einer deutschen Stichprobe hochbegabter Schüler (Preckel & Brüll, 2010).

In der IM war, wie bereits erwähnt, ein stärkeres, signifikantes Absinken in beiden Fächern zu verzeichnen. Wie zuvor für das FSK berichtet, lagen zwar auch in der IM die Mittelwerte in einem hohen Bereich und waren insgesamt linksschief verteilt, dennoch stehen sie (ebenso wie die Standardabweichungen) in Einklang mit den in anderen Studien mit identischer Operationalisierung der IM berichteten Ergebnissen (vgl. z. B. B. Spinath et al., 2006; H. S. Weber, 2012). Eine Ausnahme stellte wiederum K1 dar, in der die latenten Mittelwerte nahezu unverändert bestehen blieben, was ein Hinweis darauf sein könnte, dass in den ersten Schuljahren noch keine Abnahme in der IM zu beobachten ist und sich diese erst im weiteren Verlauf der Schulzeit manifestiert. Ein direkter Einfluss des Übergangs auf die weiterführende Schule war dabei nicht zu erkennen, vielmehr war eher von einem kontinuierlichen Absinken über den betrachteten Zeitraum hinweg auszugehen.

Dieser Befund deckt sich mit den in der Literatur berichteten Ergebnissen, wonach auch das Interesse bzw. die IM für bestimmte Schulfächer bzw. der Schule allgemein einem Abwärtstrend folgt (Bouffard et al., 2003; B. Spinath & Steinmayr, 2008; B.

¹¹ Zur Bestätigung eines BFLPE müsste das FSK positiv vom eigenen Leistungsniveau und negativ vom Leistungsniveau der Klasse beeinflusst sein. Schüler mit gleicher Fähigkeit haben demzufolge ein geringeres FSK, wenn sie in Klassen mit einem hohen Leistungsniveau platziert werden (Nagengast & Marsh, 2011).

Spinath & F. M. Spinath, 2005a). Die Ursachen dieser Entwicklung sind sicherlich vielschichtig und wurden in der Literatur in verschiedenen Ansätzen diskutiert. Neben den von Schiefele (2009) formulierten Prozessen (siehe dazu Kapitel 3.5.2) ist auch ein wachsender Einfluss der Peergruppe wahrscheinlich. Während in der Grundschule noch gute Leistungen mit sozialer Anerkennung einhergehen, verkehrt sich dies auf der Sekundarschule eher in eine soziale Ausgrenzung bei guter Schulleistung. Zudem wurden in Studien auf Grundlage von Daten aus TIMMS und PISA ebenfalls ein BFLPE für das Interesse in Mathematik aufgezeigt (Köller et al., 2000; Trautwein et al., 2006b).

Weitere wichtige Erkenntnisse, die zwar nicht in unmittelbarer Beziehung zu den Fragestellungen der Arbeit stehen, aber dennoch Aufschluss über die Struktur der motivationalen Variablen geben, liefert die Betrachtung der manifesten Interkorrelationen¹² innerhalb und zwischen den Fächern.

In diesem Punkt konnte die in der Literatur üblicherweise berichtete positive Relation zwischen dem FSK bzw. der IM in unterschiedlicher Fächer bestätigt werden. Mit Korrelationen für beide Fächer zwischen $r = 0.03$ und $r = 0.38$ im FSK bzw. $r = 0.11$ und $r = 0.35$ in der IM waren die gefundenen Zusammenhänge konsistent zu anderen Studien (B. Spinath et al., 2006; H. S. Weber, 2012). Darüber hinaus geht aus der Betrachtung der Interkorrelationen über die Zeit hinweg hervor, dass die Zusammenhänge vor allem in der Grundschulzeit im moderaten Bereich lagen und im weiteren Verlauf der Schulzeit abnahmen. Hieraus kann zunächst abgeleitet werden, dass Kinder insbesondere während ihrer frühen Schullaufbahn die Tendenz haben, ihre Fähigkeiten in den unterschiedlichen Schulfächern als ähnlich hoch einzustufen und die Inhalte der Fächer auch in ähnlicher Weise zu mögen. Dass die Korrelationen in einem moderaten Bereich liegen, spricht jedoch auch für eine fächerspezifische Differenzierung, die dann offensichtlich mit zunehmender Schulerfahrung in stärkerem Ausmaß vorgenommen wird. Dies zeigt sich in den mit wachsendem Alter sinkenden Zusammenhängen, die wiederum vorliegenden empirischen Evidenzen und theoretischen Modellannahmen entspricht (Shavelson et al., 1976; E. M. Skaalvik & Valas, 1999). Beispielsweise finden Marsh und Hau (2004) in einer Stichprobe 15-jähriger Schüler aus 26 Ländern keine signifikante Beziehung zwischen den Selbstkonzepten in unterschiedlichen Schulfächern und Steinmayr und B. Spinath (2009) berichten in einer Studie mit deutschen Oberstufenschülern (mittleres Alter: 16.94 Jahre) sogar negative Zusammenhänge zwischen FSK und Werten.¹³

¹² Die latenten Interkorrelationen stellten sich in gleicher Weise dar, lagen aber etwas höher. Die latenten Interkorrelationen wurden jedoch nicht gesondert im Ergebnisteil dargestellt, da diese nicht der Beantwortung der Fragestellungen dienten.

¹³ Die Werte Skala setzte sich bei Steinmayr und B. Spinath (2009) aus der IM sowie der wahrgenommenen Nützlichkeit und Wichtigkeit eines Schulfaches gemäß dem Erwartungs-Wert-Modell von Eccles zusammen.

Wie bereits im theoretischen Teil der Arbeit erläutert wurde, kann als Erklärung für den geringeren Zusammenhang zwischen dem mathematischen und sprachlichen Selbstkonzept im Vergleich zu den substanziellen Beziehungen zwischen den Zensuren in Mathematik und Deutsch das I/E-Modell (Marsh, 1986; vgl. Kapitel 3.5.1) herangezogen werden. Die in der vorliegenden Arbeit gezeigte Verringerung der Zusammenhänge zwischen mathematischem und verbalem FSK kann als Bestätigung der im I/E-Modell angenommenen dimensional Vergleiche, die mit steigendem Alter der Kinder häufiger werden sollen, gesehen werden (siehe dazu Dickhäuser, Seidler & Kölzer, 2005). Für die geringer werdende Beziehung zwischen der IM in unterschiedlichen Schulfächern liegt hingegen kein umfangreich geprüfetes Modell vor. In diesem Zusammenhang ist jedoch denkbar, dass hier ein Effekt der Ausdifferenzierung von Interessen mit steigendem Alter zum Tragen kommt (Schiefele, 2009) oder im Fall der IM ebenfalls dimensionale Vergleiche angestellt werden. Eventuell verändert sich auch die IM im Zuge der Veränderung der Einschätzung der eigenen Fähigkeiten (dimensionaler Vergleich). Dieser Erklärung stehen jedoch Befunde entgegen, die einen kausalen Einfluss des FSKs auf die IM nicht nachweisen konnten (Eccles & Wigfield, 1995; Obach, 2003; Wigfield et al., 1997), obwohl das FSK und die IM in einem Fach moderat bis hoch miteinander korrelieren. Dieser letzte Aspekt fand auch in der vorliegenden Studie zu sämtlichen Messzeitpunkten Bestätigung.

Klare, hypothesenkonforme Belege fanden sich in Bezug auf die interindividuelle Stabilität von FSK und IM (Hypothese 1b). Beide Konstrukte erwiesen sich in beiden Schulfächern und in allen Kohorten mit Werten zwischen $r_{tt} = 0.50$ und $r_{tt} = 0.60$ als stabil. Nur in Ausnahmen lag die Stabilität in einem niedrigeren Bereich (z. B. $r_{tt} = 0.22$, IM Deutsch, K1 Z1 nach Z3). Die Kinder behielten demzufolge ihren relativen Rangplatz innerhalb ihrer Altersgruppe im Untersuchungsverlauf in der Regel bei. Im Vergleich der manifesten zur latenten Modellierung zeigte sich eindeutig, dass die Modellierung des Messfehlers im latenten Modell zu einer höheren Stabilitätsschätzung führte. Folglich kann bei nicht latenter Modellierung in anderen Studien sogar noch eher von einer Unterschätzung der interindividuellen Stabilität im FSK und der IM ausgegangen werden. Vom gleichen Muster berichten beispielsweise auch Guay et al. (2003). Die Stabilität des FSKs und der IM lag damit in vergleichbarer Höhe mit Schätzungen aus anderen Studien (vgl. z. B. Bouffard et al., 2003; Helmke et al., 1991; Marsh et al., 1998; Marsh et al., 2007; F. M. Spinath et al., 2008a).

Neben diesem generellen Befund fanden sich auch Unterschiede zwischen den Fächer bzw. zwischen den beiden Motivationsmaßen. Zum einen war die Stabilität im FSK generell in einem höheren Bereich als in der IM. Dies bedeutet, dass in der IM häufiger Verschiebungen in der Rangreihe vorlagen als im FSK. Eventuell unterliegt die IM ei-

ner schnelleren Veränderung dahingehend, dass vor allem zu Beginn der Schulzeit von Schuljahr zu Schuljahr verschiedene Unterrichtsinhalte besonders gemocht werden. Die Stabilität nahm dann sowohl im FSK als auch der IM mit steigendem Alter zu, was dahingehend interpretiert werden kann, dass die Kinder mit steigender Schulerfahrung ihren relativen Rangplatz eher beibehalten und sich das FSK bzw. die IM auf individueller Ebene zunehmend stabilisiert.

Außerdem fiel die Stabilität in Deutsch generell höher aus als in Mathematik. Dies könnte eventuell darauf zurückzuführen sein, dass die zur Operationalisierung verwendeten Items in Deutsch ein breiteres Spektrum an Unterrichtsinhalten abdeckten und aufgrund dessen besser auf die Unterrichtsinhalte sowohl der Grund- als auch Sekundarschule passten, als dies in Mathematik der Fall war.

Im Folgenden Abschnitt wird auf die im Rahmen der Untersuchung aufgezeigten Befunde im Hinblick auf die Vorhersage schulischen Erfolgs durch allgemeine kognitive Fähigkeiten und Motivation eingegangen. Es werden zunächst die Skaleninterkorrelationen zusammenfassend diskutiert, bevor im sich anschließenden Teil die Ergebnisse der latenten Modellierung im MKSD hinzugezogen werden.

15.2 Vorhersage schulischen Erfolgs

Die Beantwortung der Frage nach den spezifischen Vorhersagebeiträgen von allgemeinen kognitiven Fähigkeiten sowie den motivationalen Merkmalen erfolgte zweischrittig. In einer ersten Betrachtung wurden die latenten Interkorrelationen herangezogen, um einen ersten Eindruck von den längsschnittlichen Beeinflussungsmustern zu erhalten.¹⁴ Die aufschlussreicheren Ergebnisse lieferte die im zweiten Schritt durchgeführte Analyse im MKSD, in der eine gleichzeitige Berücksichtigung der Stabilität schulischen Erfolgs und der Motivation, sowie deren zeitgleichen Korrelationen, möglich war. Es handelt sich hierbei also um die reinen Effekte kausaler Prädominanz zwischen schulischem Erfolg und Motivation.

Die nachfolgende zusammenfassende Diskussion und Interpretation der Befunde erfolgt analog zu diesem Vorgehen bei der Datenanalyse.

¹⁴ Die manifesten Interkorrelationen werden an dieser Stelle nicht weiter betrachtet, es soll jedoch darauf hingewiesen werden, dass diese zwar etwas niedriger ausfielen, in Richtung und Höhe aber insgesamt vergleichbar waren (siehe dazu 13.2)

15.2.1 Latente Interkorrelation der Konstrukte

Die gefundenen Zusammenhänge zwischen den allgemeinen kognitiven Fähigkeiten und SE auf der einen bzw. den motivationalen Variablen auf der anderen Seite wurden bei der Bewertung der latenten Modellierung des g -Faktors bereits eingehend diskutiert (vgl. Abschnitt 15.1.2), weshalb an dieser Stelle lediglich eine kurze Darstellung der gefundenen Korrelationen zu $Z3$ erfolgt.

Erwartungskonform lagen die Korrelationen zwischen dem latenten g -Faktor und SE zum dritten Messzeitpunkt in einem Bereich nahe 0.50 und damit in einem mit der Literatur vergleichbaren Rahmen (Gustafsson & Undheim, 1996). Zumindest für diesen latenten Faktor der allgemeinen kognitiven Fähigkeit konnte demzufolge von einer Varianzaufklärung von ca. 25% im schulischen Erfolg ausgegangen werden. Auch die gefundene etwas höhere Korrelation des g -Faktors mit der Mathematiknote korrespondiert mit vorliegenden empirischen Evidenzen (Amthauer et al., 2001).

Für die Beziehung zwischen allgemeinen kognitiven Fähigkeiten und Motivation (FSK und IM) konnten hingegen die auf Basis der Literatur erwartbaren mittleren Zusammenhänge (vgl. z. B. Helmke, 1997; Sauer & Gamsjäger, 1996) nur in K3 bestätigt werden. In K1 und K2 fand sich keinerlei Relation zwischen den Variablen, weshalb in der Schätzung innerhalb des MKSD auch keine Korrelation zwischen FSK bzw. IM und dem g -Faktor über alle Kohorten hinweg modelliert wurde. Von einer wechselseitigen Verstärkung von FSK und Intelligenz wie sie in anderen Studien (z. B. B. Spinath et al., 2006) vermutete wurde, konnte hier also nicht ausgegangen werden. Ein Alterseffekt zur Erklärung der nicht bestätigten Zusammenhänge ist im Grunde auszuschließen, da die Kinder in K1 und K2 zum dritten Messzeitpunkt bereits in der 6. bzw. siebten Klasse waren und verschiedene Studien bereits in wesentlich jüngeren Jahren auf einen Zusammenhang schließen lassen. Beispielsweise berichtet Chamorro-Premuzic et al. (2010) in der TEDS Stichprobe neunjähriger einen Zusammenhang von $r = 0.35$ zwischen einem Globalmaß für FSK und der kognitiven Fähigkeit. Es scheint sich daher eher um einen Hinweis auf eine geringere Validität des verwendeten g -Faktors zu handeln.

Für den Zusammenhang zwischen den motivationalen Konstrukten und dem SE bestätigte sich die in Hypothse 2a angenommene moderate bis hohe Relation für alle Kohorten und über die Zeitpunkte hinweg. Die zeitgleichen Korrelationen fielen dabei am höchsten aus, während die Kreuzkorrelationen zwischen den Zeitpunkten in der Regel deutlich niedriger lagen. Am geringsten waren die Zusammenhänge, wie nicht anders zu erwarten, zwischen den $Z1$ und $Z3$ Konstrukten.

Die genaue Analyse der Zusammenhänge zwischen den Konstrukten in den unterschiedlichen Kohorten und zu den drei Erhebungszeitpunkten diente letztlich einer ersten Inspektion der anzunehmenden Verursachungsrichtung zwischen Motivation und schu-

lischem Erfolg. Hiermit sollte geprüft werden, welches der in der Literatur für das FSK postulierten Modelle (*skill development*, *self-enhancement* oder *REM*) sich in einer deutschen Stichprobe unterschiedlichen Alters und unterschiedlicher Schulsysteme bestätigt. Die Berücksichtigung von IM sollte dazu beitragen, das bisherige Wissen um die Bedeutung weiterer motivationaler Merkmale neben FSK zu erweitern. Die letztliche Bewertung hinsichtlich der Verursachungsrichtung wird hier jedoch auf Basis der Ergebnisse aus dem MKSD zu ziehen sein, denn unter Anwendung dieses analytischen Vorgehens kann die absolute Höhe der interessierenden CL-Pfade beurteilt werden. Nachfolgend werden unter Bezugnahme auf die in Kapitel 13.2 berichteten latenten Interkorrelationen zunächst die Ausprägungen der zeitgleichen Korrelationen und hieran anschließend die interessierenden CL-Beziehungen zusammenfassend diskutiert.

Was den zeitgleichen Zusammenhang zwischen FSK, IM und SE angeht, wurden die in anderen Studien in unterschiedlichen Altersbereichen bereits bestätigten positiven Beziehungen auch in der vorliegenden Untersuchung gefunden. Dies bestätigt den in Hypothese 2a angenommenen Zusammenhang zwischen den Variablen und steht im Einklang mit den Ergebnissen anderer Studien. Diese belegen ebenfalls einen moderaten bis hohen Zusammenhang zwischen FSK und SE (vgl. z. B. B. M. Byrne, 1996; Guay et al., 2003; E. M. Skaalvik & Hagtvet, 1990) und einen niedrigeren, aber immer noch positiven Zusammenhang zwischen IM und SE (vgl. z. B. Gottfried, 1990; Lloyd & Barenblatt, 1984; Marsh et al., 2005b; B. Spinath et al., 2006; Steinmayr & B. Spinath, 2009). Eine Gegenüberstellung der fachspezifischen Korrelationen zeigte zudem, dass im Fach Deutsch von einem stärkeren Zusammenhang ausgegangen werden kann als in Mathematik. Dies fand sich vor allem für das FSK, bei der IM waren die Zusammenhänge eher in vergleichbarer Höhe. Auch dieser Aspekt bekräftigt die bereits angesprochene Sinnhaftigkeit einer bereichsspezifischen Operationalisierung akademischer Selbstkonzepte (Shavelson et al. (1976); siehe dazu auch Brunner et al. (2010)).

Die in Hypothese 2a postulierten höher werdenden Zusammenhänge zwischen den motivationalen Variablen und SE über den Untersuchungsverlauf hinweg, konnten hingegen nicht bestätigt werden. Vielmehr sanken die Korrelationen für beide Motivationskonstrukte (sowohl in Mathematik als auch in Deutsch), was Befunden anderer Studien, die in erster Linie für das FSK vorliegen, entgegensteht. So berichten z. B. Guay et al. (2003) in einer Längsschnittstudie einen ansteigenden Zusammenhang zwischen FSK und SE von der zweiten bis zur 6. Klasse und auch andere Studien lassen übereinstimmend auf einen höheren Zusammenhang bei steigender Schulerfahrung schließen (z. B. Chamorro-Premuzic et al., 2010; Helmke, 1997; Marsh & Craven, 2006). Auch das der IM verwandte Konstrukt *Interesse* wies in einer längsschnittlichen Analyse einen von der 7. zur 12. Klasse wachsenden Bezug zu SE auf (Köller et al., 2001). Die Autoren argumentieren dies mit einer größer werdenden Möglichkeit zu Selbstbestimmung und

eines damit verbundenen höheren Einflusses des Interesses auf die Regulation des Lernverhaltens.

Für das FSK Mathematik und Deutsch ließ sich in der vorliegenden Studie spezifischer feststellen, dass in allen Kohorten eine stärkere Abnahme der Zusammenhänge insbesondere nach dem Wechsel in die weiterführende Schule stattfand. Für die IM zeigte sich hingegen kein derart konsistentes Bild. Lediglich in K3 zeigte sich eine besonders starke Abnahme nach dem Wechsel auf die Sekundarschule. Es zeigte sich entsprechend vor allem für das FSK ein von der Literatur abweichendes Ergebnis.

Möglicherweise ist dies auch durch die in den Studien gewählte Operationalisierung der Motivation erklärbar. In vielen Studien, u.a. in der Studie von Guay et al. (2003) die aufgrund der methodischen Umsetzung von besonderem Interesse ist, wurde zur Erfassung des Selbstkonzeptes auf das *Self-Perception Profile for Children* zurückgegriffen (Harter, 1985; vgl. Muris, Meesters & Fijen, 2003), in welchem die Wahrnehmung der schulischen Kompetenz global anhand von sechs Items erfolgt (z. B. „Find, that they are very good at school“, vgl. Muris et al., 2003). Im Gegensatz dazu wurden in KoS-MoS in Anlehnung an TEDS die einzuschätzenden Kompetenzen (bzw. die Freude an der entsprechenden Tätigkeit) am fachlichen Inhalt orientiert. Demzufolge standen sie eigentlich in direkterem Bezug zu den Schulnoten. Die beobachtete sinkende Korrelation zwischen FSK bzw. IM und Schulnoten kann folglich darauf zurückzuführen sein, dass die abgefragten Inhalte vor allem zu den in der Grundschule abverlangten Kompetenzen in Beziehung stehen. Dass sich dies vor allem für das FSK zeigte, demonstriert den stärkeren Bezug zwischen FSK und SE im Vergleich zur IM. Hier wäre es denkbar, dass die Einschätzung der eigenen Fähigkeiten in enger Relation zu den konkreten Unterrichtsinhalten vorgenommen werden, wohingegen die Einschätzung, wie sehr eine bestimmte Tätigkeit gemocht wird, hiervon unabhängiger erfolgt. Auf diesen Aspekt der Operationalisierung von FSK und IM durch die verwendeten Items wird in Abschnitt 17.2 nochmal gesondert Bezug genommen. In Bezug auf die IM kann außerdem angeführt werden, dass der relativ niedrige, aber nach dem Wechsel auf die weiterführende Schule konstant bleibende Zusammenhang insofern zur Literatur passt, als sich die IM stärker auf das akademische Wahlverhalten als auf die Höhe der erzielten Leistung auswirken sollten (Eccles et al., 1983; Eccles, 2005).

Weiterhin ist zu berücksichtigen, dass die überwiegende Zahl an Studien lediglich einen Messzeitpunkt umfasst und ein direkter Vergleich der Entwicklung der Zusammenhänge somit nicht möglich war.

Für die Beantwortung der Fragestellungen ist die nun folgende Betrachtung der CL-Beziehungen, die insgesamt niedriger ausfielen als die zeitgleichen Korrelationen und mit Werten zwischen $r = 0.09$ und $r = 0.55$ in einem niedrigen bis moderat hohen Bereich lagen, von größerem Interesse. Es wird zunächst das FSK betrachtet, wobei

hier auf Basis der Interkorrelationen zum schulischen Erfolg eine erste Abschätzung hinsichtlich des Vorliegens eines *self-enhancement*, *skill-development* oder *REM* möglich ist. Für die beiden Skalen der IM wurde die Frage nach kausaler Prädominanz explorativ untersucht, da hier bisher noch keine hinreichend konsistenten empirischen Befunde vorliegen.

Durch die hier realisierte kohortenspezifische Betrachtung ergab sich für die CL-Korrelationen des FSKs ein facettenreiches Bild der Verursachungsrichtung in den unmittelbar aufeinanderfolgenden Zeitintervallen. Unter Bezugnahme auf die in Kapitel 13.2 im Detail vorgestellten Beziehungen wird nachfolgend, zunächst für das FSK Mathematik und hieran anschließend für Deutsch, eine Abschätzung der Verursachungsrichtung sowie eine zusammenfassende Einordnung in die bisherige Forschung vorgenommen. Die CL-Korrelationen für Mathematik wiesen insgesamt auf eine gewisse Altersabhängigkeit der Verursachungsrichtung hin. Während sich in der jüngsten und ältesten Kohorte im ersten Zeitintervall eine reziproke Beeinflussung abzeichnete, fand sich in K2 eher eine Verursachungsrichtung im Sinne des *self-enhancement* Modells (vorangegangenes FSK beeinflusst nachfolgenden SE). Gleiches war auch im zweiten Zeitintervall (Z2 nach Z3) für K1 zu beobachten. Dementgegen lag in den beiden älteren Kohorten in diesem Intervall (Sekundarschule) eine stärkere Beeinflussung des FSKs durch SE, also ein *skill-development* Modell, vor.

Für das FSK Deutsch lagen die Korrelationen von vorangegangenen FSK auf nachfolgenden SE bis auf eine Ausnahme (K3, Z2 nach Z3) immer etwas höher als in umgekehrter Richtung, waren aber insgesamt in ähnlicherer Höhe als im FSK Mathematik. In diesem Fall war dementsprechend ebenfalls eher von einem *self-enhancement* Modell auszugehen.

Zusammenfassend betrachtet konnte folglich sowohl für die Grundschul- als auch die Sekundarschulzeit von einer positiven Beeinflussung schulischen Erfolgs durch Fähigkeits-selbstüberzeugungen ausgegangen werden, was den Befunden zur prädiktiven Validität des FSKs entspricht (z. B. Marsh & Craven, 2006; Steinmayr & B. Spinath, 2009; H. S. Weber, 2012). Die berichteten Ergebnisse decken sich jedoch nur in Teilen mit den aus der Literatur bekannten. Zum einen wäre hier in einer frühen Phase der Schullaufbahn theoriekonform eher von einem *skill-development* Modell auszugehen gewesen (Marsh & Martin, 2011), was sich auf Basis der latenten Interkorrelationen in der vorliegenden Studie allerdings erst auf der Sekundarschule für das FSK Mathematik bestätigt hat. Für das FSK in Deutsch fand sich auf der Sekundarschule eine umgekehrte Beeinflussungsrichtung. Vor dem Hintergrund zahlreicher empirischer Studien (vgl. z. B. Marsh, 1990; Valentine et al., 2004) wäre auf der weiterführenden Schule jedoch eine reziproke Beeinflussung zu vermuten gewesen. An dieser Stelle ist für die Interpretation der Ergebnisse zur Beantwortung der kausalen Verursachungsrichtung auf Basis von ein-

fachen Korrelationen zu berücksichtigen, dass diese aufgrund der fehlenden Kontrolle von Stabilität und vorangegangener Beziehung zwischen den Variablen nur als erster Hinweis auf die dahinter liegenden Verursachungsmuster gesehen werden sollten. Bei der Modellierung im MKSD können sich diese beispielsweise dann anders darstellen, wenn die Stabilität einen großen Einfluss auf die Korrelation der Variablen hat.

Die explorative Analyse der IM ergab in beiden Fächern insgesamt niedrigere Kreuzkorrelationen als im FSK. Hier bestätigte sich für Mathematik in K1 zwischen $Z1$ und $Z2$ sowie in K2 in beiden Kreuzbeziehungen eine reziproke Beeinflussung. Die CL-Korrelation in K1 von $Z2$ nach $Z3$ fiel dann von IM auf SE höher aus. In K3 fand sich zwischen $Z1$ und $Z2$ ein stärkerer Einfluss von vorangegangener IM auf nachfolgenden SE, zwischen $Z2$ und $Z3$ kehrte sich die Beeinflussungsrichtung um. Dies bedeutet, dass vor allem von der 4. zur 6. Klasse (dies entspricht den Klassenstufen in K1 und K3 zu diesen Zeitpunkten) ein positiver Effekt einer hohen IM in der letzten Grundschulklasse auf den schulischen Erfolg in Mathematik in der 6. Klasse zu verzeichnen war.

Für das Fach Deutsch war wiederum kein zwischen den Kohorten übereinstimmendes Muster festzustellen. So fand sich in K1 und K3 jeweils von $Z1$ nach $Z2$ eine Beeinflussung der Schulnote durch die IM, wohingegen zwischen $Z2$ und $Z3$ die umgekehrte Richtung von SE auf IM stärker ausgeprägt war. In diesem Zusammenhang sollte berücksichtigt werden, dass gerade für K1 im Vergleich mit K3 die besuchte Klassenstufe zu den Messzeitpunkten weit auseinander lagen und die zuvor berichteten Einflussrichtungen dementsprechend einmal in der Grundschule und einmal in der Sekundarschule zu beobachten waren. Es lag also im Gegensatz zur IM Mathematik kein übereinstimmendes Muster von der 4. zur 6. Klasse vor. In der mittleren Kohorte fand sich dann eine gegenseitige Beeinflussung von IM und SE in sämtlichen Zeitpunkten. Auch diese Zusammenhänge bestätigen zunächst die vermutete prädiktive Validität der IM (Schiefele & Schreyer, 1994).

Insgesamt ließ diese Inspektion der latenten CL-Korrelationen noch keine abschließende Beantwortung der Fragestellung zu, denn die Richtung der Zusammenhänge stellte sich in den Kohorten und zwischen den einzelnen Konstrukten als inkonsistent heraus. Dies macht deutlich, dass eine einfache Analyse korrelativer Zusammenhänge zu kurz greift, um das komplexe Beziehungsgefüge zwischen Determinanten schulischen Erfolgs über einen längeren zeitlichen Rahmen hinweg ausreichend differenziert betrachten zu können. Aus diesem Grund wurden in der Folge komplexe strukturgleichungsanalytische Verfahren gewählt, um eine abschließende Prüfung der unter Kapitel 5 formulierten Hypothesen vorzunehmen.

15.2.2 Vorhersage schulischen Erfolgs innerhalb des MKSD

Wie im vorherigen Abschnitt bereits dargelegt, ist die reine Betrachtung der korrelativen Zusammenhänge zwischen den Variablen mit gewissen Nachteilen behaftet. Zur methodisch adäquateren Beantwortung der zentralen Fragestellung vorliegender Arbeit wurde daher eine längsschnittliche latente Modellierung im Rahmen eines MKSD realisiert. Dieses Vorgehen bietet zum einen den Vorteil einer Berücksichtigung des Messfehlers sowohl für die Einzelkonstrukte als auch über die Zeit, wodurch eine Überschätzung der Stabilität durch Methodeneffekte vermieden werden kann (Marsh & Hau, 1996). Die Messmodelle mit CU wiesen, wie aus Kapitel 12.2 hervorgeht, jeweils die beste Modellpassung auf. Ein weiterer Vorteil eines MKSD liegt in der Möglichkeit, die Zusammenhänge differenziert in den unterschiedlichen Alterskohorten zu betrachten und potenziell vorliegende Entwicklungs- und Kohorteneffekte trennen zu können. Die nachfolgende Diskussion der Befunde auf Basis des MKSD erfolgt einzeln für jedes der betrachteten motivationalen Merkmale in Mathematik und Deutsch. Der Fokus der Betrachtung wird hier auf dem FSK liegen, da die vorhandene empirische Basis für diesen Aspekt der Motivation weitaus umfangreicher als diejenige zur IM ist. Die Analyse einer weiteren motivationalen Variable diene in diesem Zusammenhang auch der Erweiterung des Wissens um die Bedeutung weiterer motivationaler Merkmale neben dem FSK.

Vor einer detaillierten Betrachtung der einzelnen Modelle kann als genereller Befund aus den MKSD Modellierungen festgehalten werden, dass die Kontrolle der Stabilität der Variablen und der zeitgleichen Korrelationen zwischen den Variablen zu einem durchaus von den latenten Korrelationen differierenden Beziehungsmuster führte. Während die latenten CL-Korrelationen mehrheitlich auf eine stärkere Beeinflussung des schulischen Erfolgs durch die Motivation (im Falle des FSKs also eher in Richtung *self-enhancement*) wiesen, fand sich in der komplexeren längsschnittlichen Modellierung eher eine Beeinflussung der Motivation durch vorangegangenen schulischen Erfolg (im Falle des FSKs also eher in Richtung *skill-development*). Des Weiteren zeigten die Ergebnisse, dass die kohortenspezifische Betrachtung der Zusammenhänge einen wichtigen Schritt bei der Beurteilung der Beziehungen darstellte: Die Full-Forward Modelle erbrachten für die Gesamtstichprobe K_AS_2 keine bedeutsamen CL-Beziehungen, während sich für die Einzelbetrachtung innerhalb der Kohorten unterschiedliche Zusammenhänge in Abhängigkeit des betrachteten Zeitintervalls ergaben. Eine einzelne Analyse der Zusammenhangsstruktur über alle Kohorten hinweg hätte demzufolge zu keinem belastbaren Ergebnis geführt.

Diese beiden Aspekte unterstreichen die Relevanz bzw. Überlegenheit einer methodisch aufwändigeren Analyse in der Beziehung zwischen schulischem Erfolg und motivatio-

nen Prädiktoren, da sich offensichtlich insbesondere an der Schnittstelle zwischen Grund- und Sekundarschule unterschiedliche Verursachungsmuster ergeben. Wie sich diese genau innerhalb der betrachteten Motivationsmaße darstellten, wird im Folgenden dargelegt.

Fähigkeitsselbstkonzept Mathematik und Deutsch

Für das FSK Mathematik konnte auf Grundlage der Modellschätzung im MKSD, also der kohortenspezifischen Modellierung der Full-Forward Modelle, ein interessanter und in dieser Form nicht erwarteter Effekt aufgezeigt werden, der in Zusammenhang mit dem Übergang auf die weiterführende Schule gesehen werden kann. In sämtlichen Kohorten fand sich ein negatives Pfadgewicht von SE auf nachfolgendes FSK in demjenigen Zeitfenster, welches den Wechsel auf die Sekundarschule beinhaltete (in K1 von Z2 nach Z3, in K2 und K3 von Z1 nach Z2). Gute Leistungen im Fach Mathematik in der Grundschule wirkten sich demnach negativ auf das FSK im ersten Schuljahr nach dem Übergang auf die Sekundarschule aus. Die gefundenen Pfadgewichte der CL-Beziehungen erreichten in den meisten Fällen keine Signifikanz, lagen aber numerisch in mit anderen Studien (und hier insbesondere Guay et al., 2003) berichteten Zusammenhängen gleich auf.

Wird der Verlauf der gegenseitigen Beeinflussung von FSK und SE näher betrachtet, findet sich zunächst in der Grundschule eine positive Beeinflussung der Fähigkeitsselbstüberzeugung im Fach Mathematik durch die erhaltene Schulnote. Dies manifestierte sich in einem höheren Pfad von SE Z1 auf FSK Z2 ($\beta = 0.11$) im Vergleich zur umgekehrten Richtung in K1 ($\beta = -0.03$). Die Kinder nutzten hier also die schulischen Leistungsrückmeldungen, um sich ein Bild ihrer eigenen Fähigkeiten zu machen (Harter, 1998, 1999). In den frühen Schuljahren ergab sich demzufolge ein *skill-development* Modell. Beim Wechsel auf die Sekundarstufe zeigte sich dann die beschriebene negative Beeinflussung des FSKs durch die vorangegangene Mathematiknote. Gerade diejenigen Kinder, die gute Noten in der Grundschule bekamen, zeigten in der Folge auf der Sekundarschule ein geringeres FSK. Im weiteren Verlauf der Sekundarschulzeit fand sich dann wiederum eine positive Beeinflussung von Schulnoten auf das FSK (*skill-development*). Hier nutzten die Kinder offensichtlich erneut die Rückmeldungen zu ihren Leistungen, um ein Bild ihrer Fähigkeiten in Mathematik zu etablieren, was sich in den Pfadschätzungen der beiden älteren Kohorten zeigte. Dieser Befund fügt sich zudem in die berichteten niedrigeren Stabilitäten im FSK im Zeitintervall des Übergangs auf die weiterführende Schule ein. Hypothese 2b konnte hiernach nur teilweise bestätigt werden: Zwar lag, wie vermutet, vor allem in der Grundschulzeit eher ein Einfluss von SE auf FSK vor, allerdings fand sich keine Bestätigung einer reziproken Beziehung auf der weiterführenden Schule.

Neben der Betrachtung der Pfadgewichte führte auch die Modellschätzung im MKSD zur Überprüfung der kausalen Prädominanz über die Kohorten hinweg¹⁵ zu einer Bestätigung der Beeinflussung von SE auf FSK. Die Fixierung dieser Pfade führte zu einer signifikanten Verschlechterung im Modell-Fit, es kann also davon ausgegangen werden, dass diese Pfade auch bedeutsam sind. Zusätzlich interessierte, ob Entwicklungs- bzw. Kohorteneffekte vorlagen. Um dies statistisch zu prüfen, wurden nur noch K1 und K3 herangezogen, da diese die Voraussetzung überlappender Zeitintervalle (4. zu 6. Klasse) erfüllten. Die Pfade dieses Zeitintervalls wurden entsprechend zwischen K1 und K3 gleichgesetzt mit dem Ergebnis, dass von einem signifikanten Entwicklungseffekt ausgegangen werden konnte, wohingegen kein Kohorteneffekt vorlag. Die festgestellte Beeinflussung des FSK durch SE zwischen dem 4. und 6. Schuljahr stellt folglich keinen, auf eine spezifische Kohorte beschränkten, Effekt dar.

Aufgrund des schlechten Fits des Messmodells des FSKs in Deutsch für K3 (vgl. Kapitel 12.2) wurde diese im MKSD nicht berücksichtigt, eine Prüfung von Kohorten- und Entwicklungseffekten war demnach für dieses Konstrukt nicht möglich.

In K1 konnte hier ein ähnliches Zusammenhangsmuster beobachtet werden, wie bereits im FSK Mathematik. Auch hier fand sich ein negatives Pfadgewicht von SE auf FSK nach dem Übergang in die Sekundarschule ($\beta = -0.13$), während in der Grundschulphase eher von einer Beeinflussung des FSKs durch SE auszugehen war ($\beta = 0.19$). In K2 fand sich schließlich ein vom FSK Mathematik abweichendes Bild. Hier lag der Pfad von FSK Z1 nach SE Z2 deutlich höher ($\beta = 0.26$ vs. 0.06). Dies bedeutet, dass die Kinder mit einer hohen Fähigkeitsselbstüberzeugung in der dritten Klasse in der fünften Klasse bessere Deutschnoten erhielten (*self-enhancement*). Von der fünften zur siebten Klasse, also im weiteren Verlauf der Sekundarschulzeit, fand sich dann wiederum ein negatives Pfadgewicht von SE auf FSK. Hier wären folglich diejenigen Kinder mit den besseren Noten zu Z2 diejenigen mit einer geringen Fähigkeitsselbsteinschätzung zu Z3.

Auf Basis der Daten aus den beiden zu betrachtenden Kohorten wäre ein Kohorteneffekt wahrscheinlicher. Dies würde bedeuten, dass es einen in den beiden Altersgruppen unterschiedlichen Verlauf in der Beziehung zwischen FSK und SE gäbe. Eine weitere statistische Überprüfung dieser Vermutung war allerdings nicht möglich. Hypothese 2c kann damit nicht als bestätigt angesehen werden.

Die zuvor berichteten Ergebnisse in Bezug auf das FSK in Mathematik und Deutsch werden nachfolgend vor dem Hintergrund der in Kapitel 3.4 dargestellten empirischen Befunde bewertet. Hierfür wird in einem ersten Schritt der Zeitraum der Grundschulzeit

¹⁵ Es handelte sich hierbei um eine Multigruppenanalyse in der spezifische Pfade fixiert wurden, vgl. 13.3.

betrachtet, hieran anschließend erfolgt die Diskussion der Verursachungsmuster am Übertritt in die Sekundarschule sowie im weiteren Verlauf der Sekundarschulzeit.

Die Ergebnisse der Analyse im MKSD belegten für den Grundschulbereich im FSK Mathematik und im FSK Deutsch in K1¹⁶ einen stärkeren Einfluss der Schulnoten auf das FSK, was inhaltlich im Sinne des *skill-development* Modells interpretiert werden kann. Dieser Befund trägt zur weiteren Klärung der Bildung von akademischen Selbstkonzepten bei, denn aus den Ausführungen zu den Ergebnissen aus Studien zum REM (vgl. Kapitel 3.4.1) ging hervor, dass noch nicht abschließend von einer Bestätigung einer reziproken Beeinflussung für jüngere Kinder ausgegangen werden kann. In Bezug auf diese Frage wurde in der Literatur unter Rückgriff auf die theoretischen Überlegungen von Harter (1999) zur Entwicklung des Selbstkonzeptes und Ergebnissen verschiedener empirischer Studien (vgl. z. B. Hattie, 1992; E. M. Skaalvik & Hagtvet, 1990) die Annahme abgeleitet, dass vor allem in der Schuleingangsphase *skill-development* Effekte vorherrschen sollten. Eine altersabhängige Veränderung in der Höhe und Richtung der Effekte konnte allerdings in einer Metaanalyse von Valentine et al. (2004) nicht bestätigt werden. Auch Guay et al. (2003) interpretieren die Befund aus ihrer Analyse im MKSD als reziproke Einflüsse sowohl in frühen Schuljahren, als auch in späteren. Im theoretischen Teil vorliegender Arbeit wurde diesbezüglich bereits argumentiert, dass auch in der letztgenannten Studie numerisch betrachtet durchaus ein stärkerer Effekt vorangegangener Leistung auf nachfolgendes Selbstkonzept vorlag. Diese Tendenz wurde zudem in einer deutschen Grundschulstichprobe bestätigt (Helmke & van Aken, 1995).

Zumindest für deutsche Grundschüler scheint demnach eher die entwicklungsorientierte Perspektive der Ansätze von Harter (1998, 1999) Gültigkeit zu besitzen. Die im Rahmen von KoSMoS gefundenen Effekte fielen dabei allerdings insgesamt niedrig aus und sind daher nur unter Vorbehalt zu interpretieren, zumal die Pfadgewichte bei Helmke und van Aken (1995) in einem wesentlich höheren Bereich lagen. Dennoch ist es durchaus plausibel anzunehmen, dass Kinder in der Schuleingangsphase vor allem Leistungsrückmeldung in Form von Zeugnisnoten nutzen, um ein Bild der eigenen Fähigkeiten zu etablieren. Dies fällt umso stärker ins Gewicht, als dass die Kinder in K1 zu diesem Zeitpunkt (zweite Klasse) ihre ersten Zeugnisse bekamen. Die Vergabe der Zeugnisse stellt hier für Kinder eine Möglichkeit zum unmittelbaren Vergleich der eigenen Fähigkeiten mit den Fähigkeiten der Klassenkameraden dar. Mit Blick auf die hohen Mittelwerte in den Schulnoten wird dieser Vergleich in der KoSMoS Stichprobe positiv ausgefallen sein. Dies bedeutet, dass die hier untersuchten Kinder wahrscheinlich zu den besseren Schülern ihrer Klasse gezählt haben (wobei hier im Grundschulbereich von einer stär-

¹⁶ K1 stellte die einzige Kohorte mit zwei Messzeitpunkten in der Grundschule dar. Diese Kohorte lieferte folglich die einzige Datenbasis für den Grundschulbereich.

keren Leistungsheterogenität im Vergleich zum Gymnasium ausgegangen werden kann; Heller, 2003), was sich dann in einem besonders positiven FSK niedergeschlagen haben könnte.

In Anbetracht der Beobachtung, dass Kinder vor allem in dieser frühen Phase ihrer Schulerfahrung ein sehr positives Bild ihrer eigenen Fähigkeiten aufweisen (vgl. z. B. D. Stipek & Iver, 1989), sind gerade auch in der Grundschule weitere Effekte denkbar, die hierauf neben Schulnoten einen Einfluss ausüben könnten. Als mögliche Erklärung kommt hier beispielsweise die sogenannte Bezugsnormorientierung (Rheinberg, 2006) der Lehrer in Frage (für einen Überblick zu dieser Forschung siehe Rheinberg & Krug, 1999). In diesem Zusammenhang konnte in verschiedenen Studien ein positiver Einfluss einer individuellen Bezugsnorm bei der Leistungsbeurteilung bestätigt werden. Bei einer *individuellen Bezugsnorm* werden die Leistungen einer Person in Relation zu deren vorangegangenen Leistungen bewertet, wohingegen eine *soziale Bezugsnorm* vorliegt, wenn die Leistungen Anderer die Vergleichsstandards definieren (Lüdtke & Köller, 2002).

Lüdtke und Köller (2002) konnten hier beispielsweise zeigen, dass durch individuelle Leistungsrückmeldungen durch die Lehrer ein zusätzlicher positiver Effekt auf das Selbstkonzept erzielt werden kann. Es wäre nun durchaus denkbar, dass gerade Grundschullehrer ihre Leistungsrückmeldungen im Rahmen einer solchen individuellen Bezugsnorm gestalten und so in besonderem Maße auf die individuellen Stärken und Schwächen eines Schülers eingehen.

Am Übergang von der Grundschule in die weiterführende Schule und im weiteren Verlauf der Sekundarschule konnte in der vorliegenden Untersuchung ein REM ebenfalls nicht bestätigt werden. Vielmehr fand sich im Zeitintervall des Wechsels in ein anderes Schulsystem ein negativer Einfluss des schulischen Erfolgs in der Grundschule auf das FSK in der weiterführenden Schule. Dies war in Mathematik in sämtlichen Kohorten und in Deutsch zumindest in K1 zu beobachten. Hier stellt sich die Frage, wie sich dieser Effekt inhaltlich interpretieren lässt.

Eine Erklärung könnten hier die bereits erwähnten Bezugsrahmeneffekte und deren Einfluss auf die Bildung des Selbstkonzeptes (vgl. z. B. Lüdtke et al., 2002) liefern (vgl. dazu ausführlich Kapitel 3.5.1). Hiermit in Zusammenhang steht sicherlich auch, dass der Übergang in die weiterführende Schule mit zahlreichen Veränderung in der sozialen wie der Lernumwelt der Kinder einhergeht, weshalb dieser in der Literatur auch als kritisches Lebensereignis (Filipp, 1995) bzw. als Entwicklungsaufgabe (Havighurst, 1972) beschrieben wurde. Gerade bei einem Wechsel auf das Gymnasium sehen sich die Kinder nicht nur mit einer Veränderung ihrer Peergruppe, sondern auch mit einer

anderen organisatorischen Strukturierung von Lernprozessen sowie der Häufigkeit, dem Modus und den Maßstäben von Leistungsrückmeldungen konfrontiert (Sirsch, 2000).

Eventuell beurteilen insbesondere in dieser Schulform die Lehrer eher innerhalb eines sozialen Bezugsrahmens, was sich direkt nach dem Übergang besonders negativ auf das schulische Selbstkonzept auswirken könnte (Lüdtke & Köller, 2002). Im Vergleich verschiedener Schultypen konnten hier schon Schwarzer et al. (1982) zeigen, dass sich bereits im Laufe des ersten Schuljahres die mittleren akademischen Selbstkonzepte von Hauptschülern und Gymnasiasten, die direkt nach dem Übergang mit einem positiveren Selbstkonzept starten, angleichen.

Im Rahmen des BFLPE haben einige Autoren (vgl. z. B. McFarland & Buehler, 1995) außerdem argumentiert, dass dieser Bezugsrahmeneffekt in erster Linie eine Folge von schlechteren Noten darstellt, die Schüler bei gleicher Leistungsfähigkeit an leistungstärkeren Schulen erhalten (Lüdtke et al., 2002). In diesem Zusammenhang hat sich zwar gezeigt, dass diese Vermutung in direkter Überprüfung des BFLPE nicht haltbar war (vgl. dazu Lüdtke et al., 2002), dennoch kann sie im Rahmen der vorliegenden Arbeit eine plausible Erklärung des negativen Effektes der letzten Grundschulnote auf das FSK nach dem Wechsel auf die weiterführende Schule (also den in den Analysen bestätigten Entwicklungseffekt) liefern.

Bei der Vergabe der Noten wählen Lehrer (und Schüler) einen klasseninternen Referenzrahmen und vergeben entsprechend der Rangordnung der erbrachten Leistungen das gesamte Notenspektrum, wobei weitgehend von einer Normalverteilung der Leistungen ausgegangen wird (Tent, 2006). Als Folge hiervon wird die gleiche Leistung in unterschiedlichen Schulkontexten unter Umständen mit vollkommen anderen Noten bewertet (Ingenkamp, 1971). Das Gymnasium kann hier nun als leistungsstarke Schulumwelt angenommen werden, denn die Schüler auf dem Gymnasium unterscheiden sich in ihren Leistungen durchaus von denen anderer Schultypen (Heller, 2003). Folglich sollten dieselben Leistungen auf dem Gymnasium mit einer schlechteren Note bewertet werden als in leistungsschwächeren Schulumwelten.

Dies bedeutet nun vor dem Hintergrund der Selektivität der KoSMoS Stichprobe hinsichtlich leistungsstarker Schüler, dass diese auf der Grundschule – die noch eine leistungsheterogene Lernumwelt darstellt (Heller, 2003) – wohl zu den Besten der Klasse gehört haben und entsprechend mit einem positiven FSK in Mathematik in die weiterführenden Schule gestartet sind. In dieser neuen Lernumwelt – hier also meist des Gymnasiums – finden sich die Kinder nun in einer leistungshomogeneren Schülerschaft wieder. Ihre Leistungen werden nun in Relation zu den Leistungen anderer, ebenfalls leistungsstarker Schüler bewertet. Hier kommt es dann wahrscheinlich erstmalig auch zu negativen Leistungsrückmeldungen, was in der Folge in einem weniger positiven FSK im Vergleich zur Grundschule münden kann. Für Kinder, die auch auf der Grundschule

schon Erfahrung mit negativen Leistungsrückmeldungen sammeln konnten, fällt dieser „Kontextschock“ eventuell weniger gravierend aus.

Kritisch anzumerken ist an dieser Stelle, dass es, wie in der vorangegangenen Diskussion bereits dargelegt, zu keinem massiven Einbruch in der Fähigkeitsselbstüberzeugung kam, da die Mittelwerte zwar absanken, aber dennoch in einem relativ hohen Bereich bestehen blieben. Die niedrigere Stabilität nach dem Übergang weist jedoch auch auf eine gewisse Veränderung in der Rangreihe hin. Um dies näher zu beleuchten, wäre in einer Folgestudie die Aufteilung der Kinder in unterschiedliche Schulformen eine wünschenswerte Erweiterung.

Der hierauf folgende positive Effekt der Note auf das FSK lässt sich wiederum inhaltlich als *skill-development* Effekt interpretieren. Hier ist denkbar, dass sich nach einer Phase der Verunsicherung in den ersten Jahren nach dem Wechsel eine erneute Etablierung des FSKs in Mathematik unter Einbezug der Leistungsrückmeldungen vollzieht. Dies steht empirisch im Gegensatz zur Annahme einer reziproken Beeinflussung von FSK und SE gerade in einem höheren Schulalter (z. B. Guay et al., 2003). In diesem Kontext wären weitere kohortenspezifische Längsschnittstudien im deutschen Raum wünschenswert, denn die meisten Befunde zur Frage der kausalen Verursachungsrichtung liegen aus amerikanischen bzw. australischen Studien vor. Im amerikanischen Schulsystem liegen jedoch weitaus weniger stark ausgeprägte Leistungsgruppierungen vor, als dies bisher in Deutschland der Fall war, was die Vergleichbarkeit der Studien einschränkt. Dieser Umstand verdeutlicht auch die Relevanz kulturvergleichender Studien, denn wenn eine Beeinflussung des Selbstkonzeptes durch die Schulumwelt vorliegt, ergeben sich hieraus auch wichtige Implikationen zur Gestaltung einer positiven schulischen Lernumwelt (siehe zu diesem Punkt z. B. Reyes, Brackett, Rivers, White & Salovey, 2012).

Abschließend bleibt noch die Frage zu klären, warum sich die Effekte nur im FSK Mathematik gezeigt haben und im FSK Deutsch in K1 und K2 differierende Effekte am Übergang in die weiterführende Schule vorlagen.

An dieser Stelle sei zunächst darauf hingewiesen, dass die Beobachtung schwächerer bzw. weniger klarer Effekte für sprachliche Fächer keine Besonderheit der vorliegenden Studie darstellt, sondern einer Vielzahl an Arbeiten zum akademischen Selbstkonzept entspricht (z. B. Marsh & Yeung, 1997; Lüdtke & Köller, 2002; Lüdtke et al., 2002). Die geringeren Effekte in der vorliegenden Arbeit gehen sicherlich teilweise auch auf die geringere Reliabilität und die weniger gute Passung des Messmodells dieser FSK Skala zurück (vgl. dazu Kapitel 11.3 sowie 12.2). Hieraus lässt sich schlussfolgern, dass mit den Items in Mathematik ähnlichere Inhalte erfasst wurden, als dies in Deutsch der Fall war. Marsh und Yeung (1997) führen in diesem Zusammenhang an, dass es im verbalen Bereich mehr Möglichkeiten zu Kompetenzerleben außerhalb der Schule gäbe,

die sich neben der Leistungsrückmeldung in Selbstkonzept niederschlugen. Weiterhin wäre es vorstellbar, dass der Mathematikunterricht den Schülern valide Informationen darüber liefert, ob eine Aufgabe richtig gelöst wurde, während der Deutschunterricht eventuell offenere Aufgabenstellungen beinhaltet. Von den Noten unabhängige soziale Vergleiche wären dann im Mathematikunterricht häufiger möglich (Lüdtke et al., 2002).

Der angesprochene Unterschied zwischen den beiden Alterskohorten kann allerdings nicht schlüssig interpretiert werden und bedarf sicherlich weiterer Studien, um methodische Artefakte ausschließen zu können, bevor eine weitergehende inhaltliche Interpretation vorgenommen werden kann.

Intrinsische Motivation Mathematik und Deutsch

Beide Variablen der IM wurden aufgrund einer bislang noch unzureichenden empirischen Ergebnislage hinsichtlich einer kausalen Verursachung explorativ untersucht, es lagen dementsprechend keine Annahmen über die Höhe und die Richtung der Zusammenhänge zu den unterschiedlichen Zeitpunkten vor. Es konnte jedoch auf Basis vorliegender empirischer Evidenzen (vgl. Kapitel 3.4.2) und der dargestellten Interkorrelationen eine positive Beziehung zwischen IM und SE über die Zeit hinweg vermutet werden.

Werden die Ergebnisse der durchgeführten Analyse zusammenfassend betrachtet, kann zunächst konstatiert werden, dass sich wiederum die kohortenspezifische Betrachtung als präzisere Analyse der Zusammenhänge erwies (vgl. dazu im Detail die Darstellung des MKSDs in Kapitel 13). Aber dennoch blieb auch unter Verwendung dieser Auswertung ein uneinheitliches Befundmuster zwischen den Kohorten bestehen, was einer belastbaren Interpretation der Befunde entgegensteht. Erschwerend kommt hinzu, dass die Pfadgewichte in der Modellierung über die drei Zeitpunkte hinweg sehr niedrig ausfielen und daher, wenn überhaupt, dann nur als Tendenzen zu interpretieren sind. Besonders ins Gewicht fiel dies in K2, da hier die Kreuzpfade in den meisten Fällen nahe Null lagen. Dennoch werden nachfolgend die Ergebnisse des MKSDs zusammenfassend dargestellt und, sofern dies sinnvoll erscheint, Interpretationsansätze vorgeschlagen. In Anbetracht der erläuterten methodischen Schwierigkeiten ist eine weitere empirische Untersuchung der Zusammenhänge in größeren Stichproben unabdingbar, um gesicherte Erkenntnisse zu erhalten.

In Bezug auf die Modellschätzung im MKSD zur Beantwortung der kausalen Prädominanz lässt sich zunächst für IM Mathematik festhalten, dass die CL-Korrelationen erwartungsgemäß niedriger lagen als im FSK, was in Einklang mit der empirischen Datenlage steht (z. B. B. Spinath et al., 2006; Toussaint, 2007; H. S. Weber, 2012).

Wie aus der Ergebnisdarstellung (vgl. Kapitel 13.3, Abbildung 13.6 und 13.8) hervorgeht lag in K1 in der Grundschulzeit ($Z1$ nach $Z2$) eine reziproke Beeinflussung, allerdings in einer eher nicht zu erwartenden Richtung, vor, denn gute Noten in der 4. Klasse gingen mit einer verminderten IM in der 6. Klasse einher, sowie eine hohe IM in der 4. Klasse mit einer besseren Leistung in der 6. Klasse. In K3 fand sich im Intervall von $Z1$ nach $Z2$ analog zum in K1 entsprechenden Schuljahr (4. zur 6. Klasse) ebenfalls diese reziproke Beeinflussung, wenngleich der negative Pfad von SE ($Z1$) auf IM ($Z2$) deutlich geringer ausfiel als in K1. Im weiteren Verlauf der Sekundarschulzeit drehte sich diese Richtung um: Hier fand sich ein negatives Pfadgewicht von IM ($Z2$) auf SE ($Z3$; $\beta = -0.12$) und ein positives Pfadgewicht von SE ($Z2$) auf IM ($Z3$; $\beta = 0.18$). Kinder mit einer hohen IM in der 6. Klasse hatten demzufolge in der 8. Klasse eher schlechtere Noten, wohingegen Kinder mit schlechteren Leistungen in der 6. Klasse in der 8. Klasse eine höhere IM aufwiesen (und umgekehrt). Dieser letzte Effekt wurde auch in K2 bestätigt, wobei hier das Pfadgewicht nur halb so hoch ausfiel. Zusammengekommen lagen die CL-Pfade also immer in einer unterschiedlichen Gerichtetheit vor, wobei das Vorzeichen von Zeitpunkt zu Zeitpunkt wechselte.

In der IM Deutsch lagen die Pfadschätzungen der CL-Beziehungen zwischen benachbarten Zeitpunkten nach Kontrolle der Stabilität und der vorangegangenen Zusammenhänge noch niedriger als in der IM Mathematik. Lediglich in K3 lagen mehrere Pfadgewichte in einem Bereich von 0.20. Auf diese Kohorte richtet sich denn auch die nachfolgende Diskussion der Befunde. Zu beiden Zeitintervallen fand sich hier eher eine Beeinflussung von IM auf nachfolgenden SE, wobei sich die Richtung der Beziehung von positiv in negativ verkehrte. Die Höhe der Schätzung war in etwa gleich ($\beta = -0.17$ und $\beta = 0.18$). Für die IM in Deutsch hatten also Kinder mit einer hohen IM in der 4. Klasse in der 6. Klasse tendenziell auch die besseren Noten. Kinder mit einer hohen IM in der 6. Klasse hatten hingegen in der 8. Klasse in der Tendenz eher die schlechteren Noten. Dieser Trend hatte sich bereits in der IM Mathematik gezeigt.

In den durchgeführten Multigruppenvergleichen führte die Gleichsetzung der CL-Pfade weder in Mathematik noch in Deutsch zu einer signifikanten Verschlechterung des Modell-Fits. Die Stärke der CL-Beziehung war entsprechend zu gering, als dass sie als empirisch bedeutsam angenommen werden konnte. Von einer weiteren Analyse hinsichtlich Kohorten- und Entwicklungseffekten wurde daher abgesehen.

An dieser Stelle bleibt die Frage zu beantworten, inwiefern sich die berichteten Ergebnisse in den bisherigen Forschungsstand integrieren lassen. Allerdings liegen bisher kaum empirische Studien vor, die sich mit dem Zusammenhang IM mit SE längsschnittlich befasst haben. Die meisten Ergebnisse liegen hier noch für das *Interesse* vor.

Für die beobachteten niedrigeren Zusammenhänge in Deutsch lassen sich hier zunächst die gleichen Argumente nennen, wie bereits für das FSK. Insofern decken sich die vorlie-

genden Ergebnisse mit Befunden anderer Studien, die ebenfalls von einem niedrigeren Zusammenhang in Deutsch berichten (vgl. z. B. H. S. Weber, 2012). Aufgrund der sehr geringen und in den Kohorten unterschiedlichen Pfadschätzungen in Deutsch fokussiert die nachfolgende Darstellung auf die IM Mathematik. Die Höhe der Pfadgewichte in Mathematik lag zumindest in der Höhe in einem Bereich, der auch in einer CL-Studie von Marsh et al. (2005b) berichtet wurde ($\beta = 0.15$). Inhaltlich schwieriger zu interpretieren sind jedoch die negativen Pfadgewichte im Verlauf der Schulzeit. Für den Zeitraum der 4. zur 6. Klasse lag hier eine positive Beeinflussung von SE durch IM vor und eine negative von SE auf IM. Dieser erste, positive Einfluss der IM auf den nachfolgenden Schulerfolg korrespondiert mit den vorhandenen empirischen Evidenzen, die einen geringen, aber konsistent positiven Einfluss der IM auf den SE vermuten lassen (z. B. Gottfried, 1990; Helmke, 1993; E. M. Skaalvik & S. Skaalvik, 2004; Steinmayr & B. Spinath, 2009). Unerwartet ist jedoch das negative Pfadgewicht von SE auf IM. Inhaltlich bedeutet dies, dass diejenigen Kinder, die in der Grundschule gute Leistungen in Mathematik haben, eine geringere IM für dieses Fach zwei Jahre später aufweisen oder umgekehrt, dass diejenigen Kinder mit schlechteren Leistungen in Mathematik in der Grundschule, in der 6. Klasse eine höhere IM aufwiesen. Dies erscheint auf den ersten Blick kontraintuitiv, da die Vermutung nahe liegt, dass sich gute Noten besonders förderlich auf das Interesse bzw. die IM auswirken sollte, diese also motivationsförderlich sind. Guten Noten wird hier eine Anreizfunktion zugesprochen, die sich dann auf die IM auswirken soll (Tent, 2006). Auf der anderen Seite sollen jedoch auch schlechte Noten dazu motivieren, Defizite zu beheben (Tent, 2006), was in der Folge eventuell zu einer höheren Anstrengungsbereitschaft führen könnte. In diesem Sinne träfe auf die vorliegende Stichprobe eher letzteres zu.

Im darauf folgenden Zeitintervall von der 6. zur 8. Klasse sind die Pfade allerdings nicht mehr in dieser Weise interpretierbar. Der negative Pfad von IM auf SE der 8. Klasse steht im Widerspruch zu der in der Literatur berichteten positiven Beeinflussung (siehe dazu insbesondere Marsh et al. (2005b), die in einer Stichprobe von Schülern der 7. Klasse positive Effekte finden). Einzig in einer Studie von F. M. Spinath et al. (2008a) wird ebenfalls eine negative Relation beider Konstrukte berichtet. Dieser wird in genannter Studie als Suppressionseffekt interpretiert. Hiervon könnte auch im Kontext der vorliegenden Arbeit ausgegangen werden, zumal nach Betrachtung der Interkorrelationen eher von einem schwach positiven Effekt auszugehen war. Dies und die Tatsache, dass die Gleichsetzung der Pfade zu keiner signifikanten Verringerung des Modell-Fits geführt hat, stellt die empirische Belastbarkeit der hier berichteten Resultate in Frage und legt die Vermutung eines methodischen Artefaktes, eventuell auch durch die starken Annahmen innerhalb eines MKSDs, nahe. Hierfür spricht auch, dass in K2 keine Effekte gefunden wurden. Von einer weiteren Interpretation wird folglich abgesehen.

15.2.3 Inkrementelle Validität des FSKs Mathematik

Als abschließende Frage innerhalb der phänotypischen Betrachtung sollte der inkrementelle Vorhersagebeitrag bzw. die kausale Prädominanz von Motivation und schulischem Erfolg unter Kontrolle allgemeiner kognitiver Fähigkeiten bewertet werden.

Wie schon an mehreren Stellen der bisherigen Diskussion erläutert, gelang die Operationalisierung der kognitiven Fähigkeiten nicht in gewünschtem Maß. Mögliche Gründe hierfür sowie die daraus abgeleiteten Konsequenzen für die vorliegende Arbeit wurden unter Abschnitt 15.1.2 dargelegt. Für die Beantwortung der unter Kapitel 5 formulierten Hypothesen 2d und 2e wurde der g -Faktor der ersten Erhebung nicht berücksichtigt, sondern lediglich der latente g -Faktor aus der dritten Erhebung herangezogen. Hierbei muss berücksichtigt werden, dass dieser sechs Jahre nach der Erfassung der Z1 Konstrukte erfolgte. Die berichteten Ergebnisse sollten daher nur unter Vorbehalt interpretiert werden, da es sich nicht um eine zeitgleiche Erfassung der Konstrukte handelte. Auf Basis der im theoretischen Teil dieser Arbeit erläuterten hohen Stabilität von Intelligenz bereits in jungen Jahren, erschien dieses Vorgehen jedoch gerechtfertigt. Vor dem Hintergrund der Ergebnisse aus der Analyse der Zusammenhänge innerhalb des MKSD wurde weiterhin allein das FSK Mathematik einer vergleichenden Betrachtung der Zusammenhänge mit und ohne Kontrolle von g unterzogen. Bei den verbleibenden Variablen waren entweder nur zwei der drei Kohorten betrachtet worden oder die CL-Beziehungen lagen in einem unbedeutenden Bereich. Im Rahmen der Modellierung des komplexen MKSD unter Kontrolle von g wurden zusätzlich zu den vorherigen Beziehungen Pfade von der allgemeinen kognitiven Fähigkeit auf den SE der drei Zeitpunkte spezifiziert. Kein Zusammenhang wurde hingegen aufgrund der nicht signifikanten, latenten Korrelationen in K1 und K2 zum FSK modelliert.

Die Ergebnisse der Modellschätzung erbrachte hier klare Belege für eine elementare Beziehung zwischen FSK und SE selbst dann, wenn der allgemein stärkste Prädiktor schulischen Erfolgs (Gustafsson & Undheim, 1996) berücksichtigt wurde. Die Evidenz dieser Beobachtung erhielt in der weiteren Modellschätzung noch stärkeres Gewicht: Es zeigte sich ein durchgängig vergleichbares Muster zur Modellierung ohne Kontrolle von g . Einzig die Pfadschätzungen der CL-Beziehungen fielen etwas niedriger aus. Nichtsdestoweniger fand sich auch in diesem Fall eine signifikante Verringerung des Modell-Fits bei Gleichsetzung der CL-Pfade zwischen den Kohorten. Die hiernach durchgeführte Überprüfung der Kohorten- und Entwicklungseffekte fiel ebenfalls vergleichbar mit der vorherigen Schätzung ohne Kontrolle allgemeiner kognitiver Fähigkeiten aus, d. h. es fand sich auch in diesem Fall ein signifikanter Entwicklungseffekt (auf dem 10% Niveau). Hypothese 2d und 2e konnten demnach für das FSK Mathematik bestätigt werden.

Die hier berichteten Ergebnisse bekräftigen einmal mehr die aus bisherigen Studien abgeleitete Annahme einer positiven Beziehung zwischen FSK und schulischem Erfolg (Lüdtke & Köller, 2002). Vorliegende Ergebnisse unterstreichen hierbei die Bedeutung des mathematischen Selbstkonzeptes aufgrund der gewählten Modellierung in einem längsschnittlichen CLP Design über mehrere Kohorten hinweg in besonderer Weise. Selbst bei Kontrolle vorangegangener schulischer Leistung, der zeitgleichen Korrelation zwischen FSK und SE, sowie des allgemein stärksten Prädiktors schulischen Erfolgs, findet sich eine bedeutsame Beziehung beider Variablen. Die Pfadgewichte lagen dabei nach Kontrolle von g meist in einem niedrigen Bereich und fielen damit geringer aus als in anderen Studien (z. B. Chamorro-Premuzic et al., 2010; B. Spinath & F. M. Spinath, 2005b; Steinmayr & B. Spinath, 2009; H. S. Weber, 2012). Auch in der Studie von Guay et al. (2003), die in Bezug auf die methodische Umsetzung die größte Vergleichbarkeit mit den vorliegenden Ergebnissen liefert, lagen die Pfadkoeffizienten etwas höher, wobei in dieser Studie keine Kontrolle der Intelligenz erfolgte und der Einfluss des FSKs in diesem Fall eventuell etwas überschätzt wird. Ohne Kontrolle allgemeiner kognitiver Fähigkeiten lagen, wie im vorherigen Abschnitt bereits beschrieben, die Pfadschätzungen in beiden Studien in vergleichbarer Höhe. Von praktischer Relevanz ist hier jedoch auch, dass der negative Pfad von SE auf nachfolgendes FSK beim Übertritt in die weiterführende Schule auch bei Kontrolle von g bestehen blieb. Eine inhaltliche Interpretation dieses Effektes wurde bereits in Abschnitt 15.2.2 vorgenommen.

Die Analyse im komplexen längsschnittlichen Modell zeigte außerdem, dass die Kontrolle der allgemeinen kognitiven Fähigkeiten auch in gewissem Ausmaß Varianz in der Stabilität der Schulnoten von $Z1$ nach $Z2$ erklärte, was sich in einer Verringerung des Stabilitätspfades manifestierte.¹⁷ Dies bedeutet, dass ein Teil der Varianz der Schulnoten, die durch die Vorleistung erklärt wurde, auf die allgemeinen kognitiven Fähigkeiten zurückging.

Nach dieser Darstellung der phänotypischen Ergebnisse der vorliegenden Arbeit werden im anschließenden Teil die ätiologischen Analysen, die in Ergänzung der phänotypischen Betrachtung durchgeführt wurden, zusammenfassend diskutiert.

¹⁷ Die größte Abnahme des Stabilitätspfades war in $K1$ zu verzeichnen, wobei hier in Rechnung gestellt werden muss, dass die Ergebnisse der Schätzung auf Schätzprobleme hinwiesen. Dieses Ergebnis sollte daher nicht überinterpretiert werden.

16 Ergebnisse der ätiologischen Betrachtung

Durch die Möglichkeit, im Rahmen der KoSMoS Studie auf genetisch sensitive Daten zurückzugreifen, konnten in der vorliegenden Untersuchung die phänotypischen Analysen durch ätiologische Analysen der Konstrukte im Quer- und Längsschnitt¹ ergänzt werden. Nachfolgend werden für jede Variable zunächst die Ergebnisse der univariaten Parameterschätzungen in Relation zu vorhandenen empirischen Evidenzen diskutiert. Hieran schließt sich die Betrachtung der multivariaten genetischen Analysen an. Die Schätzungen erfolgten zur Erhöhung der Power für die univariate Analyse auf Basis aller vorhandenen Paare zum jeweiligen Messzeitpunkt, in der multivariaten genetischen Analyse wurden die Daten aus K2 und K3 zusammengefasst (siehe dazu Kapitel 7.2.2), um eine Analyse der Ursachen von Stabilität und Veränderung im kritischen Zeitintervall des Wechsels auf die weiterführende Schule zu ermöglichen.

Sowohl bei den latenten univariaten als auch den latenten multivariaten Modellen ist in Rechnung zu stellen, dass der Modell-Fit der Schätzungen, insbesondere für den schulischen Erfolg und den g -Faktor, meist in einem grenzwertigen Bereich lag. Dies ist bei der Interpretation der Befunde dahingehen zu berücksichtigen, dass die berichteten Ergebnisse als Tendenzen verstanden werden sollten. Hier spielt auch eine Rolle, dass die zugrunde gelegten Stichproben relativ klein ausfielen in Relation zur Anzahl an zu schätzenden Parametern, was in instabilen Parameterschätzungen und großen Konfidenzintervallen resultieren kann (Kline, 2011). Inhaltlich werden nachfolgend in erster Linie Bezüge zur englischen Zwillingsstudie TEDS hergestellt, die aufgrund des ähnlichen Altersbereichs² sowie der übereinstimmenden Operationalisierung von kognitiven Fähigkeiten und Motivation eine besonders hohe Vergleichbarkeit liefert. In Bezug auf den schulischen Erfolg ist diese allerdings eingeschränkt, da in TEDS keine Schulnoten sondern Lehrerratings auf Basis des NC herangezogen wurden (siehe dazu auch Kapitel 7.1).

¹ Allgemeine kognitive Fähigkeiten fielen aus dieser Betrachtung aufgrund der fehlgeschlagenen Modellierung zu Z1 heraus. Hypothese 4a konnte demnach nicht geprüft werden.

² Das Alter lag in KoSMoS zu Z1 bei 9.06, zu Z2 bei 10.66 und zu Z3 bei 12.41. Damit sind insbesondere Z1 und Z2 mit den TEDS Daten vergleichbar.

16.1 Univariate und multivariate genetische Analyse schulischen Erfolgs

Für den schulischen Erfolg in Mathematik und Deutsch ließ bereits die Betrachtung der ICCs auf genetische Einflüsse und Einflüsse der geteilten Umwelt zu allen drei Messzeitpunkten schließen. Diese konnten auch in der Modellschätzung bestätigt werden, was den Annahmen von Hypothese 3a entspricht und im Einklang mit den Befunden anderer Studien (z. B. Walker et al., 2004) steht.

Werden die Ergebnisse detaillierter betrachtet fällt zunächst auf, dass in der Parameterschätzung der genetische Einfluss im Fach Mathematik vor allem in *Z1* höher ausfiel als im Fach Deutsch (0.78 vs. 0.45). Die Angemessenheit der Schätzung von SE Mathematik zu *Z1* ist hier allerdings kritisch zu betrachten. Zum einen würde auf Basis der ICC Schätzung, die wie in Kapitel 14.1 erläutert auf einer manifesten Skala beruhte, eher von einer Schätzung genetischer Einflüsse von ca. 50% ausgegangen werden können. Dies und die Tatsache, dass der Modell-Fit (in beiden Fächern) mit einem RMSEA von über 0.10 als nicht ausreichend bezeichnet werden kann (Kline, 2011), sprechen in *Z1* für eine Überschätzung des genetischen Effektes auf SE Mathematik in der latenten Modellierung. Zudem lagen in anderen Studien die Erblichkeitsschätzungen in sprachlichen und mathematischen Schulleistungen in etwa gleicher Höhe (vgl. z. B. Kovas et al., 2007). Aus diesem Grund wird die Parameterschätzung in Mathematik im ersten Zeitintervall nicht weiter interpretiert. Zu *Z2* und *Z3* lagen die Parameterschätzungen dann eher in einer auf Basis der ICCs erwartbaren Höhe.

Damit einher geht auch die Betrachtung einer auf den ersten Blick offensichtlichen quantitativen Veränderung der Effekte zu den drei Messzeitpunkten, die insbesondere in Mathematik aufschien aber auch in Deutsch zu beobachten war. Auch in diesem Fall sollte die Veränderung vor dem Hintergrund der oben genannten methodischen Schwierigkeiten nicht überinterpretiert werden. Dennoch ist zwischen *Z2* und *Z3* von einer gewissen Verschiebung der Parameterschätzungen zugunsten geteilter und nicht-geteilter Umwelteffekte auszugehen. Die einzige Studie mit vergleichbaren Daten stellt hier TEDS dar. Kovas et al. (2007) berichten hier von keinerlei Veränderung in den Parameterschätzungen im Alter von 7, 9 und 10 Jahren, wobei die Autoren eher von einer Erhöhung des genetischen Anteils analog zur Entwicklung der kognitiven Fähigkeiten ausgegangen waren. Diese Vermutung ließ sich auch auf Basis der deutschen Stichprobe nicht bestätigen. Die Autoren argumentieren in diesem Zusammenhang, dass die betrachtete Zeitspanne und die Abstände zwischen den Erhebungen eventuell zu kurz gewesen sei, um quantitative Veränderungen aufzeigen zu können. Was die Höhe der Parameterschätzung für den additiven genetischen Effekt angeht, lagen die Schätzun-

gen in TEDS in einem höheren Bereich als in KoSMoS. Allerdings weisen hier Walker et al. (2004) darauf hin, dass die Erblichkeit auf Basis der Lehrereinschätzung in TEDS etwas überschätzt wurde.

Ein weiterer Unterschied zwischen der deutschen und der englischen Stichprobe liegt hinsichtlich der geteilten Umwelteffekte vor: Während sich in TEDS lediglich im Fach *Naturwissenschaften* im Alter von 10 Jahren Effekte der geteilten Umwelt ergaben, weisen die Ergebnisse in der vorliegenden Studie eher auf einen substanziellen und steigenden Einfluss der geteilten Umwelt hin. Im Fach Deutsch konnte der Einfluss von C nicht ohne signifikante Verschlechterung des Modell-Fits aus der Modellschätzung entfernt werden, was deren Bedeutung weiter unterstreicht. Aber auch in Mathematik lag zu Z3 mit einer Schätzung von 32% ein bedeutsamer Einfluss geteilter Umwelteffekte vor, wenngleich sich dieser in der Modellschätzung nicht bestätigt hat. Eventuell war hier die Power der Stichprobe zu gering, um einen in der Modellschätzung signifikanten Effekt zu bestätigen (Visscher, 2004; Visscher, Gordon & Neale, 2008). Der Einfluss der nichtgeteilten Umwelt war hingegen in beiden Studien vergleichbar.

Es stellt sich nunmehr die Frage, wie sich diese Unterschiede in der Wichtigkeit geteilter Umwelteffekte zwischen den beiden Studien und zwischen den beiden Fächern erklären lassen. In diesem Zusammenhang ist es zunächst plausibel davon auszugehen, dass sich insbesondere in der Grundschulzeit Effekte der geteilten Umwelt auf schulischen Erfolg als wichtig erweisen. Neben familiären Einflüssen, die sich auf die Ähnlichkeit der Zwillinge auswirken (z. B. der SES; siehe Walker, Petrill & Plomin, 2005), besuchen die Kinder in der Regel die gleiche Schule, besuchen häufig die gleiche Klasse³ und werden oftmals von den gleichen Lehrern unterrichtet. Bei diesen schulischen Umweltfaktoren liegt es nahe, dass sie sich ebenfalls auf die Ähnlichkeit der Kinder auswirken. Aus der gleichen Überlegung heraus wurden auch von Kovas et al. (2007) signifikante c^2 Schätzungen für die jüngeren Kinder angenommen, die sich jedoch in der englischen Stichprobe nicht bestätigt haben. In der deutschen Stichprobe nahmen diese Effekte mit steigendem Alter nochmals an Bedeutung zu. Eine mögliche Erklärung der differierenden Befunde liegt in den Charakteristika der beiden Stichproben und einem sich zwischen beiden Ländern unterscheidenden Schulsystem. Hier ist zunächst die Dauer der Grundschulzeit zu nennen. Kinder im Vereinigten Königreich (UK) werden in der Regel bereits im Alter von fünf Jahren eingeschult, die Grundschulzeit beträgt dann sechs Jahre. Dies bedeutet, dass der in TEDS betrachtete Altersbereich ausschließlich die Primarschulzeit umfasst. Außerdem unterliegen die Schulen in England einem strengen Bildungsstandard, der in regelmäßigen Abständen vom *Office for Standards in Education* (OFSTED) überprüft wird (siehe dazu Huber, Späni, Schmellentin & Criblez, 2006). Zur Erreichung dieses Standards wird zudem eine hohe

³ Zu Z1 waren dies in KoSMoS 70%, vgl. Tabelle B.3.

Qualität in der Ausbildung der Lehrer angestrebt, die ebenfalls einer Beurteilung durch das OFSTED unterliegt. Vor diesem Hintergrund ist es denkbar, dass sich englische und deutsche Lehrer im Unterricht gegenüber ihren Schülern in unterschiedlicher Weise verhalten. Dieser Unterschied könnte darin bestehen, dass in UK die Lehrer eher auf die individuellen Stärken und Schwächen der Kinder eingehen, wodurch Unterschiede zwischen den Kindern begünstigt werden, welche sich dann in nichtgeteilten Umwelteffekten niederschlagen. In deutschen Schulen findet sich hingegen eventuell eher eine kollektivere Betreuung der Schüler durch die Lehrer. In diesem Fall könnte die einheitlichere Behandlung in deutschen Schulen zu einer größeren Ähnlichkeit der Kinder führen. Bezogen auf das unterschiedliche Schulsystem könnten des Weiteren auch die unterschiedlich langen Betreuungszeiten in der Schule ins Gewicht fallen. In England bleiben die Schüler in aller Regel auch am Nachmittag in der Schule und werden dort bei ihren Hausaufgaben betreut. Schulische Aktivitäten und die damit verbundenen Aufgaben sind demnach weitestgehend auch in den schulischen Alltag integriert und werden vor Ort bearbeitet. In Deutschland ist es hingegen üblich, dass die Kinder ihre Hausaufgaben nach der Schule zu Hause erledigen. Diese gemeinsame, eventuell durch die Eltern begleitete, Hausaufgabenzeit könnte für die deutschen Kinder ebenfalls einen gemeinsamen Umwelteffekt darstellen.

In diesem Punkt könnte auch die Zusammensetzung der Stichprobe, die in TEDS bevölkerungsrepräsentativ (Oliver & Plomin, 2007) und sich damit wahrscheinlich auch leistungsheterogener als in KoSMoS darstellte, eine Rolle spielen. Gerade in einer leistungsheterogenen Gruppe von Kindern wird eine individuelle Betreuung der Schüler (auch innerhalb eines Zwillingspaares) relevant. Die Beobachtung, dass sich hier in KoSMoS zu Z3, also dem Zeitpunkt innerhalb der Untersuchung zu dem alle Kinder auf der weiterführenden Schule (meist Gymnasium) waren, ein stärkerer Einfluss der geteilten Umwelt findet, passt ebenfalls in dieses Bild und kann unter Bezug auf die Diskussion der phänotypischen MKSD Ergebnisse betrachtet werden. Dort wurde bereits ausgeführt, dass sich die Lernumwelt auf deutschen Gymnasien wahrscheinlich sehr ähnlich gestaltet. Als Beispiel wurde hier eine soziale Bezugsnormorientierung der Lehrer angeführt. Ist dies zutreffend, würde sich dies genau in einem stärkeren *C* Effekt niederschlagen, denn dann würde sich die in Gymnasien bereitgestellte Lernumwelt in der Tat für alle Kinder in ähnlicher Weise gestalten, sich also auf die Ähnlichkeit der Zwillinge auswirken. Hier wäre es interessant, in weiteren Studien speziell die Unterschiede zwischen verschiedenen Schulsystemen in Deutschland zu fokussieren.

Was den Unterschied in der Höhe der Parameterschätzung in Mathematik und Deutsch angeht könnte dies zusätzlich darauf zurückzuführen sein, dass Deutsch als Muttersprache der Kinder per se einen geteilten Umwelteffekt darstellt.

Vergleichbar hoch war in beiden Studien der Einfluss der nichtgeteilten Umwelt, bei

der sich ebenfalls die Frage stellt, welche Faktoren sich in diesem Effekt wiederfinden könnten, d. h. welche Umweltfaktoren sich auf die Unähnlichkeit von Geschwistern und hier auch von EZ auswirken.

Auf der Suche nach Faktoren der nichtgeteilten Umwelt rückten zunächst Einflüsse innerhalb der Familie in den Mittelpunkt. Allerdings schlossen hier Turkheimer und Waldron (2000) in ihrer Metaanalyse, dass „measured nonshared environmental variables do not account for a substantial portion of nonshared variability“ (Turkheimer & Waldron, 2000, S. 78). Hieraus leiten Kovas et al. (2007) den Schluss ab, dass vor allem außerhalb der Familie nach möglichen nichtgeteilten Umweltfaktoren zu suchen sei. Es liegt nahe, in den Peers von Kindern einen solchen Einfluss zu vermuten, da sie eine wichtige Rolle in der individuellen Entwicklung von Kindern spielen (Harris, 1995; Iervolino et al., 2002). Weiterhin bietet sich die Wahrnehmung der Schulumwelt durch die Kinder als Einflussfaktor an (Walker & Plomin, 2006). Diese Erklärung erhält vor dem Hintergrund der Ergebnisse aus TEDS besonderes Gewicht, denn hier konnte in multivariaten Längsschnittanalysen gezeigt werden, dass nichtgeteilte Umwelteffekte vor allem spezifischer Art sind und sie zwischen unterschiedlichen Fächern nur wenig überlappen (Kovas et al., 2007). Dieses Befundmuster ist mit einem potenziellen Einfluss der Peergruppe nur schwer in Einklang zu bringen, denn dies würde bedeuten, dass sich ihr Einfluss in Mathematik und Deutsch in unterschiedlicher Weise auswirkt.

An dieser Stelle können die bisherigen Ausführungen zu den gefundenen Einflüssen auf interindividuelle Unterschiede in schulischem Erfolg auch wesentlich von den Ergebnissen der durchgeführten längsschnittlichen Analyse zu den Ursachen von Stabilität und Veränderung profitieren.⁴ Diese werden nachfolgend überblicksartig zusammengefasst und in Zusammenhang mit den dargestellten univariaten Ergebnissen interpretiert. Einschränkend sei nochmals darauf hingewiesen, dass sämtliche latenten multivariaten Modelle einen grenzwertigen Modell-Fit aufwiesen, die vorgestellten Befunde sollten daher nur als Tendenzen verstanden werden. Die Ergebnisse der durchgeführten Cholesky Zerlegung über die drei Messzeitpunkte ließen sowohl für den schulischen Erfolg in Mathematik als auch Deutsch darauf schließen, dass interindividuelle Stabilität in erster Linie durch additive genetische Effekte erklärt werden kann. Dies zeigte sich in beiden Fällen in einem starken genetischen Faktor zu $Z1$, der den größten Teil der genetischen Varianz zu $Z2$ und $Z3$ erklärte. Zu $Z2$ fand sich in beiden Fächern zusätzlich ein geringer spezifischer genetischer Effekt. Nach dem Wechsel in die weiterführende Schule kamen entsprechend noch neue genetische Einflüsse hinzu. Zu $Z3$ lag hingegen nur noch ein spezifischer Effekt im Schulfach Deutsch vor. Dies bedeutet zusammengefasst, dass genetische Effekte zu einem gewissen Teil auch Veränderung erklären

⁴ Hier sei nochmals darauf verwiesen, dass die Schätzungen aus der uni- und multivariaten Betrachtung durch die unterschiedliche Stichprobenzusammensetzung differieren können.

womit Hypothese 4b diesbezüglich als bestätigt gelten kann. Diese starke genetische Kontinuität steht im Einklang mit den aus TEDS vorliegenden Befunden (Kovas et al., 2007). Zudem konnte auf Basis der Berechnung der bivariaten Erbllichkeit festgestellt werden, dass die beobachtete phänotypische Stabilität der Schulnoten überwiegend auf additive genetische Effekte zurückzuführen war.

Interessanter ist in diesem Zusammenhang jedoch der Blick auf die Umwelteffekte. Hier fiel die geteilte Umweltvarianz generell niedriger aus als die genetische Varianz, was in großen Konfidenzintervallen und nicht signifikanten Pfadschätzungen resultierte. Für Mathematik stieg die geteilte Umweltvarianz über die Zeit hinweg an und lag zu $Z3$ bei 0.24, wobei hier eine vollständige Überlappung mit der Umweltvarianz zu $Z1$ und $Z2$ vorlag. Die geteilte Umwelt zu $Z3$ war also nicht spezifisch. In Deutsch fand sich insgesamt eine höhere geteilte Umweltvarianz, insbesondere zu $Z2$ (0.36). Zu $Z3$ lag wiederum keine spezifische geteilte Umweltvarianz vor. Dies bedeutet, dass geteilte Umwelteffekte in einem gewissen Maß sowohl Stabilität als auch Veränderung erklärten.

Für die nichtgeteilten Umwelteffekte ergab sich ein hiervon leicht differierendes Bild. Die nichtgeteilte Umwelt war vor allem zu $Z1$ und $Z2$ spezifischer Art und lag zum zweiten Messzeitpunkt, also nach dem Übergang in die weiterführende Schule am höchsten. Zudem war hier in Mathematik ein weitaus größerer Einfluss festzustellen als in Deutsch (0.40 vs. 0.22). Zu $Z3$ war der Einfluss wieder geringer und nur in Teilen spezifischer Art. Dies bedeutet, dass nichtgeteilte Umwelteinflüsse nicht nur zur Unterschiedlichkeit von Kindern, die in der gleichen Familie aufwachsen, beitragen, sondern auch, dass sie zur Veränderung von Kindern in der Zeit des Schulwechsels. Auch dieser Befund deckt sich mit den Ergebnissen aus TEDS (Kovas et al., 2007) und bestätigt die diesbezüglichen Annahmen aus Hypothese 4b.

Abschließend bleibt zu klären, ob diese Resultate die vorherigen Überlegungen zur Bedeutung der Umwelteffekte stützen oder diesen entgegenstehen. Für den geteilten Umwelteffekt kann hier zunächst festgehalten werden, dass auch in KoSMoS zu $Z2$ und $Z3$ ein substanzieller Anteil der Kinder die gleiche Klasse besuchte (56.50% respektive 55.1%) der Einfluss einer geteilten Klassenumwelt (bzw. Schulumwelt, denn die Kinder werden in den meisten Fällen zumindest die gleiche Schule besuchen) also auch als plausibel anzunehmen ist. Es scheint jedoch nicht unmittelbar erklärbar, warum sich dieser Einfluss zu $Z3$ dann nicht mehr manifestieren sollte. Es wäre höchstens vorstellbar, dass dieser Einfluss einer gleichen Klassenumwelt und des gleichen Lehrers im ersten und zweiten Jahr nach dem Wechsel auf die Sekundarstufe noch relevanter ist, denn in den meisten deutschen Schulen findet nach der sechsten Klasse (also nach $Z3$) eine Aufgliederung in einen sprachlichen und einen mathematisch-naturwissenschaftlichen Zweig bzw. in leistungsabhängige Kurse statt. Dies könnte eventuell zu einer geringeren

Konstanz der Klassenumwelt der Kinder führen, zumal in den ersten Jahren auf der weiterführenden Schule häufig mehrere Fächer von einem Lehrer unterrichtet werden. Auch die zuvor angestellte Überlegung in Bezug auf eine gemeinsame außerschulische Lernumwelt in Form der Hausaufgabenbetreuung wird wohl in höheren Schulstufen weniger relevant sein, da hier diesbezüglich eine höhere Selbstständigkeit der Kinder zu erwarten ist.

Für den nichtgeteilten Umwelteffekt kann hier der Argumentation von Kovas et al. (2007) gefolgt werden: Die nichtgeteilten Umwelteffekte waren zu einem großen Teil spezifischer Art, was einen hauptsächlichen Einfluss von Peers unwahrscheinlich erscheinen lässt. Vielmehr bekräftigen diese Ergebnisse die berichteten phänotypischen Zusammenhänge und Interpretationen. Diese nichtgeteilten Umwelteffekte, die sich speziell im Zeitintervall des Übergangs auf die weiterführende Schule finden, könnten die dort angesprochene, individuell unterschiedlich gut gelingende Anpassungsleistung der Kinder an die zum Teil gravierenden Veränderungen in der Schulumwelt widerspiegeln. Dass dies vor allem im Schulfach Mathematik zum Tragen kommt, könnte schließlich mit einer größeren Veränderung in der Unterrichtspraxis im Vergleich zu Deutsch (strengere Bewertung, häufigere Testate etc.) in Zusammenhang stehen.

16.2 Univariate genetische Analyse allgemeiner kognitiver Fähigkeiten

Für den latenten Faktor der allgemeinen kognitiven Fähigkeiten zu *Z3* erwiesen sich in der Modellschätzung vor allem geteilte und nichtgeteilte Umwelteffekte als bedeutsam (als bestfittendes Modell lag hier dementsprechend ein CE Modell vor). In der Parameterschätzung des vollständigen Modells (vgl. Abbildung 14.1) lag der additive genetische Effekt bei $a^2 = 0.26$, der Effekt der geteilten Umwelt bei $c^2 = 0.54$ und der nichtgeteilte Umwelteffekt schließlich bei $e^2 = 0.20$. Damit fiel der genetische Einfluss auf die kognitive Fähigkeitsmessung niedriger aus als auf Grundlage der in Kapitel 4.6 dargestellten Befunde für das Alter von 12 Jahren zu vermuten gewesen wäre. Hier lagen die Schätzungen für die Erblichkeit in diesem Alter meist bereits in einem Bereich von 50% (Plomin & F. M. Spinath, 2004). Auch in TEDS lagen im Alter von neun Jahren die additiven genetischen Effekte eines latent modellierten *g*-Faktors bei 43% (Greven et al., 2009). Der Anteil der geteilten Umwelt lag bei Greven et al. (2009) allerdings in ähnlicher Höhe wie in der vorliegenden Studie. Hypothese 3b kann damit nur in Teilen als bestätigt angesehen werden. Es stellt sich die Frage, wie sich diese, von der Literatur abweichenden Befunde, erklären lassen.

Hier ist zunächst festzuhalten, dass, ähnlich wie in der phänotypischen latenten Modellierung, auch in der latenten univariaten Modellierung des g -Faktors kein guter Modell-Fit erzielt werden konnte (vgl. Tabelle 14.2). Auf mögliche Gründe hierfür wurde in der Diskussion der phänotypischen Befunde bereits ausführlich eingegangen. Neben diesem methodischen Aspekt kann der unerwartet hohe Einfluss der geteilten Umwelt und der damit verbundene, geringere genetische Einfluss überdies auf inhaltlicher Ebene diskutiert werden. Dabei kann auch auf die Ergebnisse anderer Studien zurückgegriffen werden, die den g -Faktor zu $Z1$ als manifesten Faktorwert⁵ ätiologisch untersucht haben (Toussaint, 2007; Spengler, 2009). So berichtet Spengler (2009) auch für diesen Faktor einen höheren Einfluss der geteilten Umwelt und einen niedrigeren additiven genetischen Effekt (0.20) als in TEDS. Spengler (2009) argumentiert, dass dies auf potenziell vorliegendes Hilfeverhalten der Eltern speziell bei den ZZ Paaren zurückzuführen sei, welches sich dann in Kontrasteffekten⁶ niederschlagen würde. Durch die Umstellung auf eine Telefonabfrage kognitiver Fähigkeiten ist für den dritten Messzeitpunkt allerdings ein solches Hilfeverhalten der Eltern nur schwer denkbar, denn durch die Instruktion und Testdurchführung mittels eines geschulten Testleiters war die Gelegenheit für die Eltern helfend tätig zu sein, bzw. einem Mildeffekt bei Testung der eigenen Kinder zu unterliegen, stark eingeschränkt.

Viel wahrscheinlicher ist insbesondere in der dritten Erhebung ein Einfluss der in KoSMoS bereits mehrfach angesprochenen Selektivität der Stichprobe hinsichtlich des höchsten elterlichen Bildungsabschlusses. Vor diesem Hintergrund ist anzunehmen, dass die Eltern der KoSMoS Stichprobe auch hinsichtlich der Intelligenz eine Positivauswahl darstellen, diese also eher in einem oberen Bereich liegt. Wie aus anderen Studien (z. B. Jensen, 1998) hervorgeht, liegt für das Merkmal Intelligenz eine selektive Partnerwahl vor. Diese führt zu einer Erhöhung der Ähnlichkeit von ZZ, wohingegen sie auf die Ähnlichkeit der EZ keine Auswirkungen hat. Folglich führt das Vorliegen selektiver Partnerwahl zu einer Unterschätzung der Erblichkeit und einer Überschätzung der Effekte der geteilten Umwelt.

16.3 Univariante und multivariate genetische Analyse von FSK und IM

Die Ergebnisse der vorliegenden Arbeit konnten in Ergänzung der bisher noch eingeschränkten empirischen Befundlage in Bestätigung von Hypothese 3c einen klaren Nachweis der genetischen Beeinflussung beider motivationaler Konstrukte erbringen.

⁵ Die Ergebnisse einer EFA des g -Faktors ließen auf einen starken ersten unrotierten Faktor schließen, vgl. Kapitel 11.2.

⁶ Vgl. hierzu auch Kapitel 4.

Mit Blick auf die in Kapitel 14.2 berichteten Befunde hinsichtlich der ICCs und der Fit-Statistiken der latenten Modellschätzung fällt jedoch, wie bereits bei der SE in Mathematik zu *Z1*, auf, dass die auf Basis der manifesten ICCs zu vermutenden Ursachen interindividueller Unterschiede für das FSK und die IM Deutsch sehr stark von den Parameterschätzungen der latenten Modellierung abwichen. Zudem lag der Fit der univariaten Modelle hier für die meisten Kohorten zu den verschiedenen Messzeitpunkten in einem nicht mehr akzeptablen Bereich ($RMSEA > 0.08$). Dies zeigte sich in der Folge auch in der latenten Cholesky Zerlegung, in welcher der Fit mit einem RMSEA von 0.13 (für FSK und IM) höher ausfiel als in beiden mathematischen Skalen. Für diese kann in Anbetracht der Komplexität des Modells und der geringen Stichprobengröße von einem gerade noch akzeptablen Fit ausgegangen werden. Dennoch sollten auch in diesem Fall die nachfolgenden Interpretationen als erste Hinweise verstanden werden, die einer Absicherung durch weitere empirische Forschung bedürfen. Die schlechtere Passung der verbalen Motivationsskalen hatte sich hier bereits in den niedrigeren⁷ Reliabilitäten dieser Skalen sowie der weniger guten Passung der phänotypischen Modelle angedeutet. Aufgrund dieser methodischen Schwierigkeiten insbesondere für die Skalen in Deutsch wird nachfolgend von einer inhaltlichen Interpretation dieser Effekte abgesehen.

In der univariaten genetischen Analyse zu den drei Messzeitpunkten konnten im FSK und der IM Mathematik vergleichbare Muster in den Parameterschätzungen festgestellt werden. In beiden Skalen waren die ICCs der *ZZ* weniger als halb so hoch wie diejenigen der *EZ*, weshalb in beiden Fällen ein Modell mit additiven, nicht-additiven und nichtgeteilten Umwelteffekten spezifiziert wurde. Aufgrund der Konfundierung von geteilten Umwelteffekten und nicht-additiven genetischen Effekten im CTD (vgl. dazu 4.4) konnte der Einfluss der geteilten Umwelt nicht geschätzt werden.

Werden die Motivationsmaße vergleichend betrachtet, fällt zunächst auf, dass beide eine starke genetische Komponente aufwiesen. Zwischen 35% und 53% interindividueller Unterschiede wurden durch genetische Effekte, die in erster Linie nicht-additiver Art waren, erklärt. Während zu *Z1* in beiden Maßen noch additive genetische Effekte vorlagen, verschob sich dies zu *Z2* bzw. *Z3* auf nicht-additive genetische Effekte. Insgesamt wiesen die Ergebnisse der Parameterschätzung auf eine etwas stärkere genetische Beeinflussung der IM hin. Die Einflüsse der nichtgeteilten Umwelt lagen entsprechend ebenfalls zu den unterschiedlichen Messzeitpunkten in vergleichbarer Höhe. In der Modellschätzung wiesen DE Modelle die beste Anpassung an die Daten auf. Als bestfittende Modelle wurden jedoch ADE Modelle gewählt, weil sich hier die Frage stellt,

⁷ Die Reliabilitäten lagen jedoch auch in Deutsch mit > 0.60 in einem guten Bereich in Anbetracht der geringen Itemzahl.

wie sinnhaft die Annahme nicht-additiver genetischer Effekte ohne additive genetische Effekte ist (vgl. dazu 4.4).

Zur Frage der Einordnung dieser Ergebnisse in die bisherige Forschung kann einzig auf Daten aus TEDS referenziert werden, da TEDS eine der wenigen Studien darstellt, in der motivationale Variablen überhaupt auf deren Ätiologie hin untersucht wurden und zudem nur in dieser Studie eine analoge Operationalisierung der Variablen vorliegt. Ein großer Vorteil ist hier dementsprechend auch darin zu sehen, dass die Befunde zweier Studien aus unterschiedlichen Ländern vergleichend betrachtet werden können.

So finden F. M. Spinath et al. (2008a) für neunjährige Kinder⁸ ebenfalls eine starke genetische Komponente (additive und nicht-additive Effekte), diese war allerdings etwas weniger ausgeprägt als in der deutschen Stichprobe. In einer aktuellen Studien von Y. L. L. Luo et al. (2011) bestätigt sich dieser genetische Einfluss auch im Alter von 12 Jahren⁹ in ähnlicher Höhe (43%), hier wurde allerdings ein aus FSK und IM in Mathematik kombiniertes Maß verwendet. Im Unterschied zu den Ergebnissen der vorliegenden Studie fanden sich bei Y. L. L. Luo et al. (2011) im Alter von 12 Jahren aber eher additive genetische Effekte.

In Erweiterung des bisherigen Kenntnisstandes konnten in der vorliegenden Studie erstmalig auch die Ursachen von Stabilität und Veränderung über drei Messzeitpunkte hinweg analysiert werden.

Während beide Variablen in der univariaten Betrachtung weitgehend vergleichbar in ihrer Ätiologie waren, zeigten sich in der multivariaten Cholesky Zerlegung differierende Muster. Im FSK lag eine hohe genetische Kontinuität dahingehend vor, dass sowohl die additiven als auch die nicht-additiven genetischen Effekte zu $Z2$ und $Z3$ nahezu vollständig durch den ersten genetischen Faktor erklärt wurden. Genetische Effekte tragen dementsprechend in erster Linie zur interindividuellen Stabilität im FSK bei. Diese genetische Kontinuität fand sich in gleicher Weise auch in der IM, allerdings lag hier zusätzlich auch ein spezifischer, nicht-additiver genetischer Effekt zu $Z2$ vor. Für die IM erklären genetische Effekte folglich auch ein gewisses Maß an Veränderung über die Zeit. Bei den nichtgeteilten Umwelteffekten fand sich wiederum ein über beide Motivationsmaße hinweg vergleichbares Muster. Wie in Hypothese 4b angenommen, waren die nichtgeteilten Umwelteffekte in erster Linie messzeitpunktspezifisch. Zwischen $Z2$ und $Z3$ lag zusätzlich eine gewisse Überlappung der Varianz vor. Inhaltlich kann dies dahingehend interpretiert werden, dass nichtgeteilte Umwelteffekte vor allem Veränderung in beiden motivationalen Merkmalen erklären und nach dem Wechsel auf die weiterführende Schule in gewissem Maß auch zur Stabilität beitragen. Im Hinblick auf

⁸ Entspricht in etwa $Z1$ der vorliegenden Studie.

⁹ Entspricht in etwa $Z3$ der vorliegenden Studie.

die phänotypisch errechnete Stabilität zwischen den Zeitpunkten konnte auf Basis der bivariaten Erblichkeit bzw. der bivariaten nichtgeteilten Umwelt festgestellt werden, dass diese zu allen Zeitpunkten überwiegend (meist zu ca. 70%) genetisch (additive und nicht-additive) vermittelt war.

Zusammenfassend betrachtet unterstreichen die hier vorgestellten Ergebnisse die Relevanz genetischer Effekte bei der Erklärung interindividueller Unterschiede im FSK und der IM einerseits und den Einfluss der Umwelt in Form von nichtgeteilten Umwelteffekten andererseits. Dies gilt in konsistenter Weise nicht nur für unterschiedliche Geschlechter und Fähigkeiten (F. M. Spinath et al., 2008a), sondern auch in unterschiedlichen Altersbereichen und verschiedenen Nationen. Die von F. M. Spinath et al. (2008a) und Y. L. L. Luo et al. (2011) abgeleitete Forderung, in theoretischen Modellannahmen auch eine genetische Beeinflussung zu berücksichtigen, gewinnt hier durch die Bestätigung innerhalb einer deutschen Stichprobe unterschiedlichen Alters zusätzlich Gewicht.

So wird in den meisten Motivationstheorien davon ausgegangen, dass der Umwelt, z. B. dem Einfluss von Eltern und Lehrern, eine maßgebliche Bedeutung im Hinblick auf individuelle Unterschiede in motivationalen Merkmalen von Kindern zukommt und diese nicht genetisch bedingt sind (Krapp, 2005; Wigfield & Eccles, 2000). Den theoretischen Annahmen des Erwartungs-Wert-Modells zufolge sollen geteilte Umwelteinflüsse (z. B. der elterliche Erziehungsstil) das leistungsbezogene motivierte Verhalten von Kindern bedingen. Dies steht im Gegensatz zu der in verhaltensgenetischen Studien gewonnen Erkenntnis, dass nichtgeteilte Umwelteinflüsse interindividuelle Unterschiede in diesen Merkmalen erklären. In anderen Worten bedeutet dies, dass die Ähnlichkeit von Kindern in diesen Merkmalen eher ein Resultat ihrer genetischen Ähnlichkeit darstellt und, dass sich der Einfluss von Eltern oder Lehrern individuell spezifisch auf die Kinder auswirkt (Y. L. L. Luo et al., 2011). Worin genau dieser nichtgeteilte Umwelteinfluss besteht, ist, wie bereits beim SE, ungeklärt. Hier wäre denkbar, dass Kinder, die in der gleichen Familie aufwachsen, diese Familienumwelt in unterschiedlicher Art und Weise erleben (Plomin & D. Daniels, 1987). Möglicherweise werden vergangene Erfahrungen und wahrgenommene Erwartungen wichtiger Bezugspersonen von den Kindern unterschiedlich interpretiert. Mit Blick auf die Ursachen von Stabilität und Veränderung ist auch denkbar, dass gerade beim Wechsel auf die weiterführende Schule neue Umwelteinflüsse salient werden, die von Kindern in unterschiedlicher Weise wahrgenommen werden und sich dann spezifisch auf das FSK und die IM der Kinder auswirken. Beispielsweise könnte sich hier eine unterschiedlich gute Anpassung an die neuen Lernbedingungen wiederfinden. Diese könnte sich dann im weiteren Verlauf der Sekundarschulzeit zu einem gewissen Grad auf die Stabilität im FSK und der IM auswirken.

An dieser Stelle sei jedoch auch kritisch angemerkt, dass diese Ergebnisse nicht zwangsläufig mit einer Negierung sämtlicher geteilter Umwelteinflüsse gleichzusetzen sind. Die in den ICCs beobachtete geringe Ähnlichkeit der ZZ (aus der letztlich die Schätzung von ADE Modellen abgeleitet wurde), könnte teilweise auch auf andere Effekte zurückzuführen sein, was wiederum eine Überschätzung der genetischen Anteile zur Folge hätte. Zum einen könnten Kontrasteffekte vorgelegen haben. Die Zwillinge könnten sich bei der Einschätzung ihrer Fähigkeiten also nicht mit der Population, sondern mit ihrem Geschwister verglichen haben und dabei die Unterschiede zwischeneinander überschätzt haben. Saudino et al. (1995) konnten hier nachweisen, dass Kontrasteffekte bei ZZ stärker ins Gewicht fallen als bei EZ, was sich dann in einer Überschätzung genetischer Einflüsse manifestiert. Es wäre auch möglich, dass die EEA hinsichtlich der motivationalen Merkmale nicht haltbar ist und sich die Umwelten von EZ in stärkerem Maße ähneln als die Umwelten der ZZ. Hier wäre vorstellbar, dass sich EZ in ihren Fähigkeiten ähnlicher sind als ZZ und in Folge dessen auch eine ähnlichere Lernumwelt bereitgestellt bekommen. Nichtsdestotrotz sind die Ergebnisse dieser und anderer Zwillingsstudien als ein starker Beleg genetischer Einflüsse auf das FSK und die IM zu werten.

16.4 Verhaltensgenetische Cross-Lagged Analyse

In der abschließenden genetischen Betrachtung des in der phänotypischen Analyse gefundenen Entwicklungseffekts vom SE in der vierten Klasse auf das FSK in der sechsten Klasse wurde auf das von Y. L. L. Luo et al. (2010) vorgeschlagene Vorgehen zurückgegriffen. Übereinstimmend mit der in Hypothese 5a¹⁰ formulierten Annahme, ergab sich eine überwiegend genetische Vermittlung des CL-Pfades ($\beta = 0.20$). Nahezu 70% des Entwicklungseffekts wurden durch genetische Effekte erklärt. Die verbleibende Varianz wurde durch nichtgeteilte Umwelteffekte vermittelt. Dies steht in Einklang mit den von Y. L. L. Luo et al. (2011) berichteten Befunden auf Basis der TEDS Daten. Auch dort fand sich eine hauptsächlich genetische Vermittlung des Zusammenhangs von SE auf FSK.

Hier bleibt zu erörtern, in welcher Art und Weise die ätiologischen Analysen die phänotypisch gefundenen Zusammenhänge hinsichtlich Kohorten- und Entwicklungseffekten ergänzen können. Kohorteneffekte spiegeln in erster Linie spezifische Aspekte einer bestimmten Umwelt wider, in der eine bestimmte Alterskohorte aufwächst (Lohaus et al., 2010). Diese haben sich in der phänotypischen Analyse nicht bestätigt. Übereinstimmend zeigte sich in der ätiologischen Analyse, dass Umwelteffekte nur einen kleinen

¹⁰ Hypothese 5b wurde nach den Befunden aus der phänotypischen Analyse nicht geprüft.

Teil des Effektes von SE auf FSK erklären. Ein allgemeiner Entwicklungstrend unabhängig von der jeweiligen Kohorte steht daher in Einklang mit den gefundenen hohen genetischen Effekten. Y. L. L. Luo et al. (2011) argumentieren an dieser Stelle, dass die gefundene genetische Überlappung Pleiotropie impliziere und daher vermutlich Gene zu finden seien, die mit dem SE im Alter von 9 Jahren und dem FSK im Alter von 12 Jahren assoziiert sind. Diese Interpretation stellt jedoch nur eine mögliche Erklärung für das Zustandekommen genetischer Korrelationen dar (Johnson et al., 2011). Eine genetische Korrelation kann nicht nur dann zustande kommen, wenn ein gleiches Set an Genen zwei Merkmale beeinflusst, sondern wäre auch dann zu beobachten, wenn ein genetisch beeinflusstes Merkmal A ein anderes (nicht genetisch bedingtes) Merkmal B beeinflusst oder vice versa. Möglich ist auch, dass ein drittes, genetisch beeinflusstes Merkmal C (in diesem Zusammenhang beispielsweise Intelligenz) sowohl A als auch B beeinflusst und die Korrelation zwischen beiden durch die geteilten genetischen Effekte mit C verursacht wird. Welche Prozesse sich hinter der beobachteten genetischen Korrelation verbergen ist demnach nicht definitiv zu klären. Die Ergebnisse weisen jedoch darauf hin, dass umweltbedingte Effekte nur zu einem kleinen Teil den Effekt des SE auf FSK in dieser Altersklasse bestimmen. Es ist jedoch nicht auszuschließen, dass Effekte der geteilten oder nichtgeteilten Umwelt im Rahmen einer Gen-Umwelt Interaktion zu einem früheren oder späteren Zeitpunkt in der Lebensspanne Einfluss auf den SE oder das FSK nehmen.

Nach dieser Zusammenschau der Ergebnisse der durchgeführten Analysen und deren Einordnung in den breiteren Forschungskontext wird abschließend eine zusammenfassende Interpretation vorgenommen und sich hieraus ergebende Implikationen abgeleitet. Letztlich erfolgt eine kritische Diskussion der Grenzen vorliegender Arbeit und ein Ausblick auf künftige Forschungsfragen, deren Beantwortung ein vertiefendes Verständnis der Zusammenhänge erlauben würden.

17 Fazit

17.1 Abschließende Zusammenfassung und Implikationen der Studie

Die vorliegende Studie verfolgte mehrere Ziele. Zum einen wurde auf phänotypischer Ebene der Zusammenhang zwischen schulischem Erfolg und den motivationalen Merkmalen FSK und IM über einen Zeitraum von sechs Jahren betrachtet, um deren kausale Prädominanz zu bewerten. Das verwendete MKSD hat sich dabei als adäquate Methode erwiesen, unterschiedliche Effekte des Zusammenspiels am Übertritt in die weiterführende Schule, eines für Kinder einschneidenden Entwicklungserlebnisses, zu erkennen. Zum anderen sollte überprüft werden, inwiefern FSK und IM auch inkrementell über allgemeine kognitive Fähigkeiten hinaus zur Vorhersage schulischer Leistung beitragen können. Ergänzt wurden diese phänotypischen Betrachtungen durch ätiologische Analysen, die zum Ziel hatten, das Wissen um Ursachen interindividueller Unterschiede sowie von Stabilität und Veränderung von klassischerweise als umweltbedingt angenommenen Variablen zu erweitern.

Resümierend kann an dieser Stelle zunächst festgehalten werden, dass die aussagekräftigsten Ergebnisse für das FSK in Mathematik gefunden wurden, was im Einklang mit anderen Studien steht und die Wichtigkeit einer fachspezifischen Beurteilung von Zusammenhängen unterstreicht. Für die IM fanden sich hingegen bei Berücksichtigung der Stabilität und der zeitgleichen Korrelationen keine bedeutsamen Beziehungen über die Zeit. Diese geringere Bedeutung der IM hatte sich dabei bereits in anderen Studien gezeigt (Steinmayr & B. Spinath, 2009; H. S. Weber, 2012). Dennoch ist nicht auszuschließen, dass es sich dabei um ein altersspezifisches Muster handelt und dieses bei längerer zeitlichen Betrachtung der Sekundarschulzeit eine Veränderung erfährt (Steinmayr & B. Spinath, 2009).

Für das FSK in Mathematik offenbarte die Analyse im MKSD ein anderes Beziehungsmuster als in vielen vorliegenden Studien, die eher für eine reziproke Beeinflussung zwischen FSK und SE sprachen (z. B. Guay et al., 2003; Marsh et al., 1999). Wie in Entwicklungstheorien des FSKs vermutet, nutzen Kinder in der Grundschule die Leistungsrückmeldungen in Form von Noten, um ein Bild ihrer eigenen Fähigkeiten

zu etablieren. Die Befunde der vorliegenden Studie legen nun nahe, dass es mit dem Übertritt in die weiterführende Schule zumindest für leistungsstarke Grundschüler, die auf das Gymnasium wechseln, zu Bezugsrahmeneffekten kommt, die das FSK in Mathematik beeinflussen. Dies wurde auch an der geringeren Stabilität des FSKs im Zeitintervall des Wechsels auf die weiterführende Schule deutlich. Zudem konnte gezeigt werden, dass dieser Effekt nicht kohortenspezifisch war, sondern als Entwicklungseffekt interpretiert werden kann. Gerade für längsschnittliche Studien, die die Entwicklung in einem spezifischen Merkmal fokussieren, können verhaltensgenetische Studien nun durch die Abschätzung genetischer und Umwelteinflüsse maßgebliche Informationen über die Ätiologie dieser Merkmale bereitstellen (Leve, Harold, Ge, Neiderhiser & Patterson, 2010). Der hier erbrachte Nachweis einer genetischen Beeinflussung motivationaler Merkmale in unterschiedlichen Altersbereichen bekräftigt die Forderung von beispielsweise F. M. Spinath et al. (2008a), die theoretischen Modellvorstellungen des Erwartungs-Wert-Modells von Eccles und Kollegen (Eccles et al., 1983) um eine genetische Komponente zu erweitern. Als weitere wichtige Erkenntnis in Bezug auf die Modellvorstellungen kann hier angeführt werden, dass sich Umwelteinflüsse nicht wie postuliert in Form geteilter Umwelt manifestieren, sondern nichtgeteilter Art sind. Letztlich erbrachte die multivariate genetische Analyse, dass insbesondere genetische Einflüsse zur Stabilität interindividueller Unterschiede beitragen, während nichtgeteilte Umwelteinflüsse insbesondere in der Zeit des Übergangs zur weiterführenden Schule Veränderung erklären.

Abschließend bleibt zu bewerten, welche Implikationen aus den Ergebnissen der vorliegenden Studie abzuleiten sind und in welcher Art und Weise die Erkenntnisse ätiologischer Analysen zur Erweiterung des bestehenden Wissens beitragen können. Motivationale Voraussetzungen von Schülern gelten in der aktuellen pädagogisch-psychologischen Forschung als wichtige Personenmerkmale, die leistungsbezogenes Verhalten erklären und vorhersagen können. Eine Vielzahl empirischer Studien belegt deren prognostische Relevanz für schulischen Erfolg auch über Merkmale wie Intelligenz und Persönlichkeit hinaus (z. B. Helmke & van Aken, 1995; B. Spinath, Freudenthaler & Neubauer, 2010; Steinmayr, Bipp & Spinath, 2011; Valentine et al., 2004). Besonderes Forschungsinteresse lag hier auf akademischen Selbstkonzepten, die sich in vielfältiger Weise auf das Erleben und Verhalten von Personen in Leistungssituationen auswirken sollen (Wigfield & Karpathian, 1991). Vor dem Hintergrund empirischer Studien, die beispielsweise eine verminderte Anstrengungsbereitschaft in Mathematik u.a. mit der (mathematischen) Selbstwirksamkeit in Verbindung bringen können (Martin, Anderson, Bobis, Way & Vellar, 2012), wird die unmittelbare praktische Relevanz der Frage nach Steigerung bzw. Aufrechterhaltung eines positiven Selbstkonzeptes und der Implementierung einer solchen in den Lehr- und Lernalltag von Kindern offenkundig. Auch wenn die Effektgrößen nach Kontrolle der Stabilität

(die im Falle der Noten auch das Vorwissen umfasst) und der zeitgleichen Korrelation in einem niedrigeren Bereich liegen, kommt ihnen doch erhebliche praktische Relevanz zu. Aus den empirischen Befunden einer positiven Beziehung zwischen Noten und Selbstkonzept, die sich auch in der vorliegenden Studie auf korrelativer Ebene bestätigt hat, sollte jedoch nicht geschlossen werden, dass die Vermittlung eines hohen, wenn auch unrealistischen akademischen Selbstkonzeptes besonders sinnvoll sei. Ganz im Gegenteil kann die Vermittlung und Aufrechterhaltung eines überschätzten FSKs sogar negative Auswirkungen auf den Lernerfolg von Schülern ausüben. Haben die Kinder die Überzeugung, die notwendigen Fähigkeiten bereits zu besitzen und schulischen Erfolg auch bei geringer Anstrengung als selbstverständlich erachten, kann dies in einer geringeren Lernbereitschaft münden und letztlich auf der weiterführenden Schule negative Konsequenzen haben (Helmke, 1998). Diese Überlegung gewinnt vor dem Hintergrund des hier aufgezeigten Entwicklungsverlaufs von der Grundschule in die weiterführende Schule bei Kindern, die in der Grundschule sehr leistungsstark waren, zusätzlich an Bedeutung.

Besonders erfolgreich scheinen Kinder mit einem leicht überschätzenden FSK zu sein. Bei diesen ist es vorstellbar, dass sie sich auch solche Aufgaben zutrauen, die sie bei vollständig realistischer Einschätzung nicht in Angriff genommen hätten und dabei auch die notwendige Anstrengungsbereitschaft an den Tag legen (Helmke, 1998). An dieser Stelle sollte zudem nicht unberücksichtigt bleiben, dass die Ausdifferenzierung von Fähigkeitsbereichen, die vor allem während der Sekundarschulzeit im Rahmen dimensionaler Vergleiche evident wird, einen wichtigen Schritt in der Identitätsentwicklung darstellt (Havighurst, 1972). Akademische Selbstkonzepte geben einer Person Auskunft über die eigenen Stärken und Schwächen und tragen so zur Wahl geeigneter (im Sinne von passender) Umwelten bei (Möller & Köller, 2004). Entwicklungspsychologisch gesehen ist eine Anpassung von Fähigkeitsselbstüberzeugungen gerade nach dem Wechsel auf die weiterführende Schule also durchaus funktional. Als Konsequenz der empirischen Evidenzen leitet B. Spinath (2005) die praktische Empfehlung ab, den Kindern ein aufgabenbezogenes Feedback bereitzustellen, das es ihnen ermöglicht, realistische Selbsteinschätzungen der eigenen Fähigkeiten zu entwickeln und eine optimale Passung zwischen Fähigkeit und Aufgabenschwierigkeit zu wählen. In diesem Zusammenhang scheint auch die Wichtigkeit einer individuellen Bezugsnormorientierung der Lehrer wieder auf, die ebenfalls zu einer realistischen Einschätzung beitragen kann (Lüdtke & Köller, 2002).

Letztlich bleibt die Frage, wie die angenommene, relativ leichte Beeinflussbarkeit motivationaler Variablen vor dem Hintergrund der verhaltensgenetischen Analysen, die eine starke genetische Komponente nahelegen, zu bewerten ist. Ein Aspekt bei der korrekten Interpretation verhaltensgenetischer Studien liegt im Verständnis der Erblieh-

keit als Populationsstatistik, die letztlich keine Aussage über die Veränderbarkeit eines Merkmals erlaubt. Trotz einer hohen Erbllichkeit können sich die mittleren Ausprägungen im SE, dem FSK und der IM über die Zeit oder nach gezielten Interventionen verändern, wobei anzunehmen ist, dass die relative Position eines Kindes innerhalb der Alterskohorte recht stabil bleibt (Plomin et al., 2008). Den Einfluss der Umwelt betreffend konnten in der vorliegenden Untersuchung für SE und Motivation unterschiedliche Beiträge aufgezeigt werden. Für den schulischen Erfolg ergab die Analyse, dass geteilte Umwelteffekte in der deutschen Stichprobe interindividuelle Unterschiede erklären, während diese in der englischen Stichprobe unbedeutend waren. Mögliche Gründe für diesen kulturellen Effekt in der Lernumwelt der Kinder wurden in der Diskussion ausführlich erörtert. Korrespondierend mit TEDS fanden sich auch in KoSMoS für die motivationalen Merkmale ausschließlich nichtgeteilte Umwelteffekte, deren Relevanz die Ergebnisse der längsschnittlichen Analysen vorliegender Arbeit zusätzlich unterstreichen.

In der Umwelt liegt schließlich auch der Schlüssel zu einer Verknüpfung von phänotypischen und ätiologischen Analysen. Viele Studien konnten beispielsweise einen positiven Effekt von Interventionen, die in erster Linie auf eine Veränderung in den Umweltbedingungen zielen, für Kinder, Eltern und Erwachsene in unterschiedlichen Bereichen belegen (Leve et al., 2010). Dennoch zeigt sich immer wieder, dass viele Kinder von solchen Interventionen nicht profitieren oder der positive Effekt nicht überdauernd ist, Kinder also interindividuell unterschiedlich auf die bereitgestellte Umwelt reagieren.

Hier liefern genetisch sensitive Daten einen guten Ansatzpunkt, differentielle Umwelteffekte besser zu verstehen und für die Praxis einen Nutzen zu ziehen. Wie bereits im theoretischen Teil der Arbeit ausgeführt, kann es dabei nicht mehr um ein entweder Anlage *oder* Umwelt gehen, sondern um ein Zusammenspiel von beidem in Form von Gen-Umwelt Korrelationen und Gen-Umwelt Interaktionen (z. B. Freese, 2008; Neiderhiser et al., 2004; Jaffee & Price, 2007; Rutter, Pickles, Murray & Eaves, 2001; Turkheimer & Waldron, 2000). Indem verhaltensgenetische Studien Informationen darüber bereitstellen, wie spezifische Umweltvariablen mit Genen interagieren bzw. kovariieren und auf eine bestimmte interessierende Variable wirken, können von dieser Forschung wesentliche Impulse beispielsweise für die Interventions- und Präventionsforschung ausgehen (Leve et al., 2010). So konnten beispielsweise r_{GE} Studien aufzeigen, dass genetisch beeinflusste Merkmale des Kindes zu harschen und kontrollierenden Reaktionen der Eltern führen, welche sich dann als Risikofaktor bei der Entwicklung von Verhaltensauffälligkeiten der Kinder erwiesen haben (vgl. z. B. Leve et al., 2010). Erziehungstrainings können dann in der Folge durch Komponenten ergänzt werden, die den Eltern alternative Reaktionen auf diese Merkmale ermöglichen. Ein anderes Beispiel stellt eine Studie von Shanahan, Vaisey, Erickson und Smolen (2008) dar, die zeigen konnten,

dass das soziale Kapital einer Familie¹ den Zusammenhang zwischen einer bestimmten genetischen Disposition, die als Risikofaktor auffälligen Verhaltens zählt (DRD2 Variante), und der Fortsetzung des Schulbesuchs vermittelt. Darüber hinaus können längsschnittliche verhaltensgenetische Studien genutzt werden, um Entwicklungsperioden zu identifizieren, in denen ein spezifischer Umwelteffekt am stärksten mit einer bestimmten Outcome-Variable in Beziehung steht (Leve et al., 2010).

Das kritische Moment bei der Verknüpfung von Verhaltensgenetik und phänotypischer Forschung stellt daher die Identifikation und adäquate Erfassung von relevanten Umweltmerkmalen dar, die so direkt und objektiv wie möglich erfolgen sollte um empirisch belastbare Ergebnisse zu erzielen (siehe zu diesem Punkt z. B. Moffitt, Caspi & Rutter, 2005; Rutter et al., 2001; Shanahan, Hofer & Shanahan, 2003). Auch bei der Frage nach möglichen Interventionen zur Steigerung bzw. Aufrechterhaltung von motivationalen Merkmalen kann die Verhaltensgenetik wichtige Anhaltspunkte liefern. Hierfür wird künftig die genaue Erfassung von Umweltvariablen im schulischen und familiären Bereich ausschlaggebend sein. Zu denken ist hier beispielsweise an bestimmte Aspekte des Erziehungsverhaltens oder des Involvements der Eltern (dies sind insbesondere kontrollierende Verhaltensweisen), die sich bereits als wichtige Prädiktoren schulischen Erfolgs erwiesen haben (z. B. Karbach, Gottschling, Spengler, Hegewald & Spinath, 2013) oder Indikatoren des sozialen Kapitals (Shanahan et al., 2008).

Nachfolgend werden schließlich die methodischen und studienspezifischen Grenzen der vorliegenden Studie betrachtet, vor deren Hintergrund die Ergebnisse der vorliegenden Arbeit zu bewerten sind, bevor ein Ausblick auf weitere Forschungsfragen gegeben wird.

17.2 Limitationen

Eine erste Limitation der Generalisierbarkeit der Befunde vorliegender Untersuchung ergibt sich aus der verwendeten Stichprobe, die in mehrerer Hinsicht selektiv war.

Durch die Notwendigkeit der Rekrutierung einer Zwillingsstichprobe lag bereits zum ersten Messzeitpunkt eine Einschränkung hinsichtlich des SES (höchster elterlicher Bildungsabschluss) vor. Hierin spiegelt sich der, auch in anderen Studien beobachtbare Effekt wider, dass vor allem obere Bildungsschichten eine Affinität zur Teilnahme an wissenschaftlichen Studien aufweisen und aus diesem Grund auch eher bereit sind, an solchen aufwändigen Studien teilzunehmen. Damit einher geht auch eine Selektivität

¹ Dieses setzt sich zusammen aus einem hohen SES, einem hohen Involvement der Eltern und einer guten Schule.

hinsichtlich der Bearbeitungsmotivation. Das in KoSMoS eingesetzte Fragebogen-set war sehr umfangreich, so dass die Familien mit einem erheblichen Zeitaufwand rechnen mussten. Dies war besonders in der ersten Erhebung relevant, denn hier fungierten die Eltern für die kognitive Testung auch als Testleiter. In Bezug auf die Fortsetzung der Teilnahme an KoMSoS ergab sich im weiteren Studienverlauf eine weitere Selektion hinsichtlich des SES, was sich auch in der Häufigkeit der gewählten Schultypen nach der Grundschule zeigte (vgl. dazu Kapitel 7). Hier bestätigte sich, dass die Kinder mehrheitlich auf das Gymnasium wechselten, was im Einklang mit der zuvor bereits angesprochenen Beobachtungen steht, dass vor allem Kinder aus Familien mit hohem SES eine Gymnasialempfehlung erhalten (Ditton, 2010b). Trotz dieser Selektivität liefern die Ergebnisse wichtige Erkenntnisse, gerade bezogen auf einen höheren Fähigkeitsbereich von Kindern.

Eine weiterer wichtiger Aspekt im Rahmen von Längsschnittstudien betrifft die Frage nach einem systematischen Unterschied zwischen Teilnehmern, die auch über einen längeren Zeitraum an einer Studie partizipieren und solchen, die nach einer gewissen Zeit abbrechen (z. B. Baltes, Reese & Lipsitt, 1980). Hiermit in Zusammenhang steht auch die Frage nach der Zulässigkeit einer, wie auch in der vorliegenden Untersuchung vorgenommenen, Ersetzung fehlender Daten, welche durchaus auch als kritisch angesehen werden kann. Hier ist zu beurteilen, was schwerer wiegt: Eine Schätzung fehlender Daten oder ein selektiver Stichprobenausfall über den zeitlichen Verlauf hinweg. Die Exaktheit der Datenersetzung die durch neue statistische Verfahren erreicht werden kann, spricht allerdings eindeutig dafür, dass ein größerer Fehler bei Nichtersetzung fehlender Daten begangen werden würde (Allison, 2002; Davey, Shanahan & Schafer, 2001). Aus diesem Grund wurden auch in dieser Arbeit fehlende Werte ersetzt, wenn mindestens aus zwei der drei Erhebungen Daten der Kinder vorlagen. Dennoch sind die Ergebnisse insbesondere zum ersten Messzeitpunkt nur unter Vorbehalt zu interpretieren, da hier viele Indikatoren schulischen Erfolgs mittels ML-Schätzung ersetzt wurden. Die hohe Stabilität der Noten (Grube & Hasselhorn, 2006; Tent, 2006) rechtfertigt dieses Vorgehen jedoch und lässt eine systematische Ergebnisverzerrung unwahrscheinlich erscheinen.

Schließlich wiesen auch die hohen Mittelwerte in den Noten auf eine Selektivität hinsichtlich des Fähigkeitsniveaus der Kinder hin, die meist gute bis sehr gute Noten in der Grundschule erhielten. Dies führt letztlich auch zu einer Varianzeinschränkung des Kriteriums und geringeren Zusammenhängen zwischen Prädiktor und Kriterium (Asendorpf, 2004).²

Ein letzter Punkt hinsichtlich der Stichprobenauswahl der kritisch beleuchtet werden

² Dies zeigte sich auch in höheren Pfadschätzungen im zweiten Untersuchungsintervall, in dem die Varianz des Kriteriums etwas höher lag.

sollte, stellt die Analyse von Zwillingen und die Frage nach der Generalisierbarkeit auf die Population dar. So ist es durchaus vorstellbar, dass sich Zwillinge gerade in einem solchen Merkmal wie dem FSK eher mit ihrem Zwillingsgeschwister vergleichen als mit anderen Kindern, denn Zwillinge stellen füreinander die perfekte Vergleichsgruppe dar. In diesem Sinne könnten hier auch dimensionale Vergleiche zwischen den Geschwistern auftreten. Für die phänotypische Schätzung hätte dies eine Varianzeinschränkung zur Folge, was sich wiederum in niedrigeren Pfadschätzungen niederschlagen könnte. Dennoch konnte in verschiedenen Studien die Übertragbarkeit von Ergebnisse aus Zwillingstudien auf die Population bestätigt werden (z. B. Posthuma et al., 2000). Zusätzlich wurde in der Modellschätzung um die Abhängigkeit der Stichprobe korrigiert (siehe dazu 9). Im Endeffekt resultiert aus dieser Korrektur ein niedrigerer Modell-Fit und schlägt sich darin nieder, dass die Pfadschätzungen weniger schnell signifikant werden. Diese Korrektur um Abhängigkeit der Daten ist vor allem auch dann kritisch zu hinterfragen, wenn es um den Vergleich verschiedener Submodelle geht, die aufgrund der Korrektur die Signifikanzgrenze nur knapp unter- bzw. überschreiten. Liegen größere Stichproben vor, wäre hier eine zufällige Auswahl eines Zwilling pro Paar als alternatives Vorgehen vorzuziehen.

Neben diesen stichprobenspezifischen Einschränkungen sind zusätzlich verschiedene methodische Aspekte zu diskutieren.

Diese beziehen sich zunächst auf die Operationalisierung der Konstrukte, die Passung der latenten Messmodelle und die Schätzung des MKSD. Für die Skalen der motivationalen Variablen kann hier zunächst festgehalten werden, dass diese, in Anbetracht der Tatsache, dass lediglich drei Items erfasst wurden, eine gute Reliabilität in unterschiedlichen Altersbereichen aufwiesen. Dennoch sollte nicht unbeachtet bleiben, dass die Reliabilität von der ersten zur dritten Erhebung abnahm. In der latenten Modellierung über drei Messzeitpunkte erwiesen sich die Skalen jedoch als messinvariant, d. h. es kann davon ausgegangen werden, dass zu den drei Zeitpunkten Gleiches erfasst wurde. Allerdings wurde auch in der latenten Modellierung deutlich, dass vor allem in den jüngeren Kohorten ein guter Modell-Fit erreicht werden konnte, was für eine gewisse Altersabhängigkeit der Schätzung spricht. Eine Erklärung hierfür liegt in der Formulierung der Items, die nahe am Unterrichtsinhalt der Grundschule orientiert war. Eine geringere Passung der Items für ältere Kinder wurde auch darin deutlich, dass Item 2 aus dem Messmodell zu Z3 entfernt werden musste um ein fittendes Modell zu erhalten. Außerdem konnte für das FSK in Deutsch für K3 kein adäquater Modell-Fit erreicht werden, was zum Ausschluss dieser Kohorte aus den weiteren Analysen führte. Für künftige Studien wäre daher zu überlegen, in älteren Kohorten eine Anpassung der Items vorzunehmen. Hiermit würde allerdings einhergehen, dass die Messungen in verschiedenen Altersbereichen nicht mehr direkt miteinander vergleichbar wären.

Auch in Bezug auf den SE ergeben sich Einschränkungen in der Modellierung. Die Angemessenheit der Operationalisierung schulischen Erfolgs durch Schulnoten per se wurde im theoretischen Teil der Arbeit bereits angesprochen. Diese bezieht sich in erster Linie auf einen Mangel an verbindlichen Richtlinien zur Notenvergabe und darauf, dass Noten innerhalb eines klassenspezifischen Bezugsrahmens vergeben werden. Beides führt zu einer verringerten Vergleichbarkeit der Schulnoten (Tent, 2006). Die nachgewiesene hohe Korrelation zwischen Noten und Leistungstestergebnissen spricht an dieser Stelle jedoch für eine im Allgemeinen gute Einschätzung der Leistungen durch die Lehrer (Tent, 2006), was den Einsatz von Noten in Schulstudien rechtfertigt, zumal sie das realere Kriterium schulischen Erfolg darstellen (Steinmayr & B. Spinath, 2009). Neben diesem allgemeinen Einwand kann auch die latente Modellierung durch nur zwei Indikatoren hinterfragt werden. Diese wurde gewählt, um zeitliche Parallelität der Messungen zu gewährleisten und damit den geforderten Voraussetzungen eines *cross-lagged panel* Designs gerecht zu werden (Marsh et al., 1998). Hierdurch lagen fachspezifisch jeweils nur die Halbjahres- und Jahresendnote des entsprechenden Messzeitpunktes vor. Diesem Vorgehen steht die Empfehlung entgegen, mindestens drei Indikatoren pro latenter Variable zu spezifizieren (Kline, 2011). In diesem Zusammenhang wäre zu überlegen, ob die Halbjahresnote des Folgeschuljahres noch in die Modellierung aufgenommen werden sollte. Gerade vor dem Hintergrund der Fragestellung vorliegender Studie, die auch den Übergang auf die weiterführende Schule und die damit einhergehenden Veränderungen fokussierte, erschien dies allerdings wenig sinnvoll.

Auf die Schwierigkeiten, die mit der Erfassung der kognitiven Fähigkeiten einhergingen wurde bereits ausführlich eingegangen, weshalb dieser Aspekt an dieser Stelle nicht weiter ausgeführt wird.

Die in der phänotypischen Schätzung bereits aufscheinende geringere Reliabilität der Skalen und die teilweise geringeren Ladungen auf die latenten Faktoren schlugen sich auch in der ätiologischen latenten Schätzung nieder, was zu Folge hatte, dass der Fit der univariaten latenten Modelle nicht ausreichend war, und zum anderen die Parameterschätzungen stark von den manifesten Skalen abwichen. Die teilweise geringen Stichprobengrößen können hier ebenfalls eine Rolle gespielt haben.

Eine generellere Frage stellt sich in Bezug auf die Erfassung der Merkmale mittels Fragebögen. Diese stellen eine ökonomische Variante dar, Daten in großen Stichproben zu erheben. Kritisch diskutiert werden müssen an dieser Stelle allerdings Antworttendenzen im Sinne sozialer Erwünschtheit. So wäre es durchaus vorstellbar, dass die Kinder sich in besonderer Weise als fähig und interessiert dargestellt haben, weil sie dies als wünschenswerte Verhaltensweise interpretieren. Die Analyse der deskriptiven Kennwerte ließ hier jedoch nicht auf eine gravierende Antwortverzerrung schließen,

denn die Verteilungsmuster waren insgesamt schlüssig und fielen konsistent zu anderen Studien aus.

Schließlich bleibt auch die Ermittlung der Zygocie mittels Fragebogen zwangsläufig hinter einem DNA Test zurück. Die Verteilung der Zygocie entsprach jedoch größtenteils derjenigen in der Population und der eingesetzte Fragebogen erwies sich an anderer Stelle als sehr zuverlässig was die Bestimmung der Zygocie anbelangt (Price et al., 2000b). Aus diesen Gründen kann hier von einer ausreichenden Sicherheit bezüglich der Zygocieinschätzung ausgegangen werden.

Abschließend sollen noch einige Punkte betrachtet werden, die das Zwillingsdesign als Methode der Erblichkeitsschätzung betreffen. Zunächst gilt bei der Interpretation von Erblichkeitsschätzungen allgemein, dass es sich dabei, wie bereits erwähnt, um eine Gruppenstatistik handelt, die den Anteil genetischer Varianz an der Gesamtvarianz angibt. Dies bedeutet, dass jede Schätzung von Erbe und Umwelt eine Momentaufnahme der betrachteten Stichprobe zur betrachteten Zeit darstellt, womit sich eine unmittelbare Übertragung auf andere Altersbereiche und Kulturen verbietet. Aus dieser statistischen Betrachtung geht auch hervor, dass die Schätzung der Erblichkeit auch immer davon abhängt, wie homogen sich die Umwelt gestaltet. Wären alle Kinder einer exakt gleichen Umwelt ausgesetzt, läge die Erblichkeit bei 100%. Ein hoher genetischer Anteil in einem Merkmal kann daher auch bedeuten, dass die Umweltvarianz in der Stichprobe gering war. An dieser Stelle sollte zudem bedacht werden, dass sich bei einem Merkmal wie dem FSK oder der IM eventuell gerade in jüngeren Stichproben schnelle Veränderungen ergeben könnten und ein zeitlicher Abstand von zwei Jahren zwischen jeder Erhebung unter Umständen zu groß war, um diese zu erfassen.

Eine bereits an verschiedenen Stellen angesprochene Schwierigkeit des CTDs liegt in der Konfundierung von geteilten Umwelteffekte und nicht-additiven genetischen Effekten. Hier bedarf es weiterer Verwandtschaftsgruppen um beide Effekte simultan schätzen zu können (vgl. z. B. Hahn et al., 2012). Zusätzlich ist hier zu bedenken, dass nicht signifikante Effekte der geteilten Umwelt nicht mit deren Negierung gleichzusetzen sind. Poweranalysen haben allerdings deutlich gemacht, dass für einen entsprechenden Nachweis sehr große Stichprobengrößen benötigt werden (Visscher, 2004; Visscher et al., 2008). Daneben weisen Johnson et al. (2011) darauf hin, dass beispielsweise das Vorliegen einer genetischen Korrelation nicht notwendigerweise auf Pleiotropie zurückzuführen ist und dass weiterhin nicht eindeutig geklärt ist, was unter Dominanzeffekten genau zu verstehen ist. Schließlich sind auch die Umwelteffekte nur unter Vorbehalt als reine Umwelteffekte anzunehmen, da in ihnen auch r_{GE} bzw. GxU Interaktionen enthalten sein können.

Die inhaltliche Interpretation der Studienergebnisse kann vor dem Hintergrund der dargestellten methodischen Grenzen nur vorläufig sein und bedarf weiterer, vorzugsweise längsschnittlicher und genetisch sensitiver Studien, die eine breite Erfassung familiärer und individueller Determinanten schulischen Erfolgs bieten.

17.3 Ausblick

Aus der vorangegangenen Darstellung der Ergebnisse, Implikationen und Limitationen dieser Arbeit werden abschließend Forschungsfragen abgeleitet, die eine sinnvolle Ergänzung des bisherigen Wissens und der Ergebnisse dieser Studie darstellen.

Aus den Ergebnissen der phänotypischen und ätiologischen Analysen ergibt sich zunächst, dass eine vergleichende Betrachtung unterschiedlicher Schulsysteme auf Länderebene bei der Identifikation von Unterrichtsmerkmalen mit Einfluss auf die schulische Lern- und Leistungsentwicklung wichtige Hinweise liefern kann.

Dies wird einerseits dadurch deutlich, dass sich in der deutschen Stichprobe andere Verursachungsmuster zwischen Motivation und SE ergeben haben, als dies in amerikanischen Studien der Fall war. Andererseits haben sich auch in der ätiologischen Analyse in der englischen und deutschen Stichprobe unterschiedliche Einflüsse auf interindividuelle Unterschiede in der Schulleistung ergeben. Hier wäre ein gezielter Vergleich von Lernumwelten in verschiedenen Kulturkreisen und deren Auswirkungen auf die Lern- und Leistungsentwicklung eine sinnvolle Ergänzung um der Frage nach besonders förderlichen Umweltbedingungen für eine gelingende Schullaufbahn weiter auf den Grund zu gehen. Dass sich bestimmte Aspekte der Schulumwelt auf die Leistung der Kinder auswirken, wurde dabei bereits verschiedentlich nachgewiesen (z. B. Reyes et al., 2012). Sinnvoll wäre hier auch eine Erweiterung der betrachteten Prädiktoren insbesondere auf familiäre Faktoren. In diesem Zusammenhang wird auf theoretischer Ebene von einem komplexen Beziehungsgefüge zwischen familiären und individuellen Bedingungsfaktoren ausgegangen (Helmke & Weinert, 1997) und verschiedene Studien konnten hier bereits erste Hinweise liefern, dass insbesondere kontrollierendes Verhalten der Eltern negativ mit der Schulleistung assoziiert ist (Karbach et al., 2013; Levpušček & Zupančič, 2009; Seginer, 2006). Allerdings gilt es hier, in längsschnittlichen Studien auch das Verhalten der Kinder in die Betrachtung einzubeziehen, denn es ist durchaus vorstellbar, dass kontrollierendes bzw. eingreifendes Elternverhalten nur die Reaktion auf das Verhalten des Kindes darstellt. Die Berücksichtigung vieler Einflussfaktoren und die Modellierung komplexer Zusammenhänge (insbesondere auf Basis von Strukturgleichungsanalysen) über mehrere Zeitpunkte hinweg bedarf jedoch in jedem Fall umfangreicher und qualitativ hochwertiger Datensets.

Weiterhin wurde deutlich, dass der betrachtete Altersbereich der Kinder von entscheidender Bedeutung ist. In Folgeuntersuchungen läge ein wichtiger Forschungsauftrag vor allem darin, insbesondere Übergänge, die wichtige Einschnitte im Leben von Kindern darstellen, einer genaueren Analyse zu unterziehen. Hier wären kürzere Zeitintervalle zwischen den einzelnen Testungen oder eine engmaschige Erfassung relevanter Prädiktoren im Schuljahr nach dem Übertritt in die weiterführende Schule sinnvolle Erweiterungen der vorliegenden Studie.

Auch ein gezielter Vergleich der Entwicklungsverläufe in unterschiedlichen Schulformen stellt eine interessante Erweiterung dar. Wie aus der Ergebnisdarstellung und der Interpretation bereits hervorging, beziehen sich die Aussagen der vorliegenden Studie hinsichtlich des Zusammenhangs von SE und mathematischen FSK in erster Linie auf Gymnasialschüler und Kinder im oberen Leistungsbereich. An dieser Stelle könnte eine Untersuchung mit leistungsheterogeneren Stichproben bzw. unterschiedlicher Schultypen wesentlich dazu beitragen, das vorhandene Wissen bezüglich der Einflussvariablen auf schulischen Erfolg zu erweitern. Einige Studien geben in diesem Zusammenhang erste Hinweise auf eine unterschiedliche Wirkung beispielsweise von Bezugsrahmeneffekten in unterschiedlichen Lernumwelten (Makel et al., 2012; Preckel & Brüll, 2010).

Vor diesem Hintergrund erhält auch die Frage, wie eine optimale Gestaltung des Übergangs in die weiterführende Schule aussehen müsste, besonderes Gewicht und auch in der aktuellen politischen Diskussion um die Sinnhaftigkeit einer verlängerten Grundschulzeit bzw. einer sogenannten Gemeinschaftsschule wurde diese Frage aufgeworfen. In einer Reanalyse der Daten der ELEMENT Studie, in der die schulischen Leistungen von Kindern mit vierjähriger und sechsjähriger Grundschulzeit verglichen wurden, sehen Baumert et al. (2009) entgegen der ursprünglichen Interpretation (Lehmann & Lenkeit, 2008) keinen Nachteil in einer verlängerten Grundschulzeit. Vielmehr würden sich leistungsstarke Kinder auch auf der Grundschule durchsetzen und dies läge unter anderem an individuellen Lernvoraussetzungen der Kinder und deren familiärem Hintergrund. Die bisherige Datenlage zu diesem Thema sei hier aber noch nicht ausreichend, um einen für alle Kinder gültigen Schluss ziehen zu können (Baumert et al., 2009). Hieraus wird deutlich, dass das Gelingen des Übergangs in die Sekundarstufe auch an individuelle Voraussetzungen bzw. Persönlichkeitsmerkmale des Kindes geknüpft ist. Die Suche nach möglichen Moderatoren, die beispielsweise im Bereich von Copingressourcen, Selbstwirksamkeitserwartungen oder Attributionen zu vermuten wären, würde an dieser Stelle auch Anhaltspunkte für gezielte Fördermaßnahmen liefern können.

Schließlich wurde deutlich, dass die Kontrolle der Intelligenz bzw. der schulischen Vorleistung unabdingbar ist, um die Effekte der Prädiktoren auf die schulische Leistung fundiert beantworten zu können. Hier wäre in Ergänzung der Erfassung der allgemei-

nen kognitiven Fähigkeiten auch an die Berücksichtigung des Arbeitsgedächtnisses als weitere Variable der kognitiven Eingangsvoraussetzungen zu denken (Alloway et al., 2005; H. S. Weber, 2012).

Für den Bereich der verhaltensgenetischen Forschung wurde bereits herausgestellt, dass eine exakte Erfassung von Umweltvariablen in Kombination mit molekulargenetischen Analysen sowie die Erweiterung auf umfangreichere und heterogenere Stichproben den nächsten Schritt in der Erforschung von Ursachen interindividueller Unterschiede in motivationalen Merkmalen darstellen müssten. Der Wert verhaltensgenetischer Längsschnittstudien ist hier insbesondere darin zu sehen, dass auch die Ursachen von Stabilität und Veränderung betrachtet werden können und sich daraus praktische Implikationen ableiten lassen. Darüber hinaus würden zukünftige Studien auch von einer Erweiterung um andere Verwandtschaftsgruppen (z. B. Geschwister oder adoptierte Kinder) und einer Analyse von Gen-Umwelt Korrelationen sowie Gen-Umwelt Interaktionen profitieren.

Messmodell allgemeine kognitive Fähigkeiten

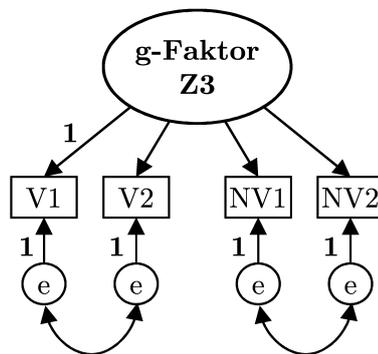


Abbildung 1.3. Messmodell der allgemeinen kognitiven Fähigkeiten. V1 = Wortanalogien; V2 = Allgemeinwissen; NV1 = Figurenklassifikation; NV2 = Figurenanalogien.

Messmodell Schulerfolg

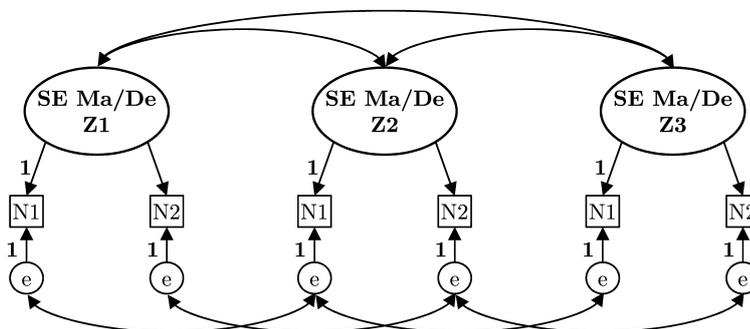


Abbildung 1.4. Messmodell Schulerfolg. N1 = Halbjahresnote; N2 = Jahresnote.

Multivariates verhaltensgenetisches Modell

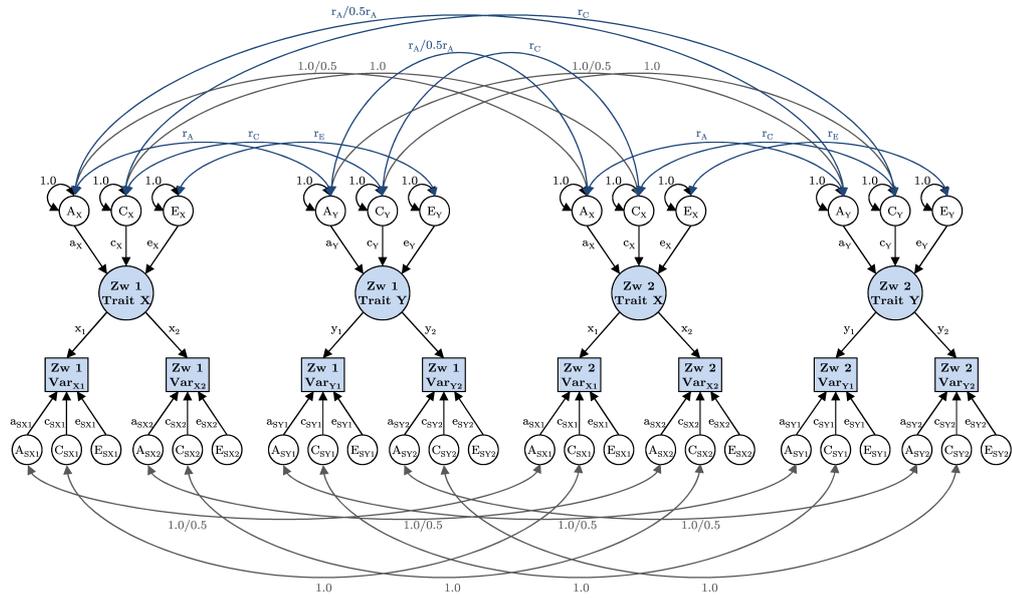


Abbildung 1.5. Multivariates verhaltensgenetisches Modell. Eigene Darstellung in Anlehnung an Plomin und F. M. Spinath (2002).

B Tabellenanhang

Daumenregeln zur Beurteilung des Modell-Fits

Tabelle B.1

Gängige „Daumenregeln“ zur Bewertung der Modellgüte

Fit-Index	Guter Fit	Akzeptabler Fit
χ^2	$0 \leq \chi^2 \leq 2df$	$2df < \chi^2 \leq 3df$
p -Wert	$0.05 < p \leq 1.0$	$0.01 \leq p \leq 0.05$
χ^2/df	$0 \leq \chi^2/df \leq 2$	$2 < \chi^2/df \leq 3$
<i>RMSEA</i>	$0 \leq RMSEA \leq 0.05$	$0.05 < RMSEA \leq 0.08$
Konfidenzintervall (<i>CI</i>)	Nahe <i>RMSEA</i> , linke Grenze des <i>CI</i> = 0	Nahe <i>RMSEA</i>
<i>SRMR</i>		$0 \leq SRMR \leq 0.08$
<i>CFI</i>	$0.97 \leq CFI \leq 1.0$	$0.95 \leq CFI < 0.97$
<i>TLI</i>	$0.95 \leq TLI \leq 1.0$	$0.90 \leq TLI < 0.95$
<i>AIC</i>	Kleiner als <i>ACI</i> des Vergleichsmodell	

Anmerkungen. In Anlehnung an Schermelleh-Engel et al. (2003) und Weiber und Mühlhaus (2010). *df*=degrees of freedom; *RMSEA*=Root Mean Square Error of Approximation; *SRMR*=Standardized Root Mean Square Residual; *CFI*=Comparative Fit Index; *TLI*=Tucker-Lewis-Index; *AIC*=Akaike Information Criterion.

Notation in Pfadmodellen

Tabelle B.2

Notation der Variablen im Pfadmodell der SEM

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 Messvariablen x
ζ	Fehlerterm der latenten endogenen Variable
β	Pfadgewicht zwischen exogener und endogener Variable
λ	Faktorladungen
γ	Pfadgewicht zwischen exogener und endogener Variable
\rightarrow	Regressionsgewicht
\leftrightarrow	Kovarianz bzw. Korrelation
	Latente Variable
	Manifeste Variable
	Fehlerterm

Anmerkungen. In Anlehnung an Weiber und Mühlhaus (2010, S.40).

Gesamtstichprobe KoSMoS über den Untersuchungsverlauf hinweg

Tabelle B.3
Zusammensetzung der Gesamtstichprobe K_AS

	Gesamte Stichprobe			Kohorte 1			Kohorte 2			Kohorte 3		
	Z 1	Z 2	Z 3	Z 1	Z 2	Z 3	Z 1	Z 2	Z 3	Z 1	Z 2	Z 3
Anzahl	776	504	392	233	150	115	294	194	145	248	163	132
Alter	9.06	10.66	12.41	8.27	9.83	11.60	9.08	10.60	12.46	9.84	11.47	13.17
Geschlecht												
Männlich	383	240	188	117	70	58	151	99	74	115	72	55
Weiblich	393	264	204	116	80	57	143	95	71	133	91	77
Klassenstufe				233								
2. Klasse	233											
3. Klasse	295	10			10		294					
4. Klasse	248	151	1		139	1		11		248	1	
5. Klasse		178	9			9		177			31	
6. Klasse		33	117			105		2	11		131	1
7. Klasse		131	158						130			28
8. Klasse			103						2			101
ohne Angabe		1	4		1			4	2			2
Schultyp												
Grundschule	776	160	1	233	149	1	294	11		248	1	
Gymnasium		210	235			76		109	86		102	73
Gesamtschule		32	32			9		15	15		17	8
Realschule		88	88			19		51	37		38	32
Hauptschule		11	15			9		8			3	6
Sonstiges		2									2	
ohne Angabe		1	21		1	1			7			13
Gleiche Klasse												
Absolut	540	287	216	147	97	70	214	110	78	179	80	68
in Prozent	69.6	56.5	55.1	63.1	64.7	60.9	72.5	56.4	53.8	72.2	49.1	51.5

Zusätzliche Tabellen der phänotypischen Ergebnisse

Tabelle B.4
Faktorladungen der Messmodelle in den Kohorten

		Zeitpunkt 1		Zeitpunkt 2		Zeitpunkt 3	
		N1	N2	N1	N2	N1	N2
SE_Ma	K1_AS_2	0.95	0.99	0.87	0.79	0.94	0.86
	K2_AS_2	0.94	0.85	0.86	0.87	0.92	0.86
	K3_AS_2	0.84	0.87	0.83	0.85	0.89	0.86
SE_De	K1_AS_2	0.97	0.98	0.91	0.91	0.92	0.91
	K2_AS_2	0.92	0.99	0.91	0.82	0.86	0.84
	K3_AS_2	0.84	0.99	0.90	0.82	0.89	0.82

Anmerkungen. N(K1) = 170; N(K2) = 215; N(K3) = 190; -SE = Schulerfolg; Ma = Mathematik; De = Deutsch; N1 = Halbjahresnote; N2 = Jahresnote.

		Zeitpunkt 3			
		NV1	NV2	V1	V2
g-Faktor	K1_AS_2	0.62	0.59	0.56	0.58
	K2_AS_2	0.34	0.39	0.69	0.39
	K3_AS_2	0.47	0.60	0.56	0.61

Anmerkungen. N(K1) = 170; N(K2) = 215; N(K3) = 190; NV1 = Figurenklassifikation; NV2 = Figurenanalogien; V1 = Wortanalogien; V2 = Allgemeinwissen.

		Zeitpunkt 1			Zeitpunkt 2			Zeitpunkt 3		
		I1	I2	I3	I1	I2	I3	I1	I2	I3
FSK_Ma	K1_AS_2	0.41	0.80	0.73	0.57	0.88	0.79	0.58	0.78	0.76
	K2_AS_2	0.60	0.80	0.76	0.50	0.95	0.71	0.82	0.81	0.84
	K3_AS_2	0.56	0.87	0.74	0.71	0.92	0.81	0.71	0.75	0.82
FSK_De	K1_AS_2	0.56	0.65	0.70	0.62	0.79	0.58	0.55	0.79	0.60
	K2_AS_2	0.64	0.70	0.72	0.62	0.58	0.55	0.52	0.91	0.61
	K3_AS_2	0.69	0.65	0.74	0.62	0.86	0.69	0.68	0.58	0.68
IM_Ma	K1_AS_2	0.46	0.63	0.99	0.46	0.70	0.77	0.47	0.83	0.84
	K2_AS_2	0.53	0.87	0.73	0.49	0.84	0.79	0.57	0.94	0.81
	K3_AS_2	0.41	0.84	0.76	0.45	0.92	0.86	0.60	0.89	0.87
IM_De	K1_AS_2	0.51	0.70	0.67	0.62	0.80	0.68	0.37		0.53
	K2_AS_2	0.57	0.77	0.73	0.54	0.93	0.52	0.48		0.56
	K3_AS_2	0.53	0.80	0.65	0.41	0.89	0.67	0.53		0.68

Anmerkungen. N(K1_AS_2) = 170; N(K2_AS_2) = 215; N(K3_AS_2) = 190; FSK = Fähigkeitsselbstkonzept, IM = Intrinsische Motivation; Ma = Mathematik; De = Deutsch; I1 = Item 1; I2 = Item 2; I3 = Item 3. Item 2 in FSK_De zu Zeitpunkt 3 entfernt.

Tabelle B.5
Prüfung der Messinvarianz Zeit – Schulerfolg Mathematik und Deutsch

SE_Ma	Modell	Vergleich	χ^2	df	χ^2/df	CFI	TLI	SRMR	AIC	Δdf	$\Delta\chi^2(\text{TRd})$
K1_AS_2	M1		7.21*	2	3.61	0.99	0.92	0.01	1432.72		
	M2	M1	18.08**	4	4.52	0.97	0.89	0.07	1438.24	2	11.69**
	M4	M1	8.34*	3	2.78	0.99	0.95	0.05	1432.29	1	1.19
	M3	M2	5.93	6	0.99	1.00	1.00	0.06	2283.48	2	0.78
K2_AS_2	M1		1.47	2	0.74	1.00	1.01	0.01	2286.63		
	M2	M1	4.96	4	1.24	1.00	0.99	0.05	2285.95	2	3.61
	M3	M2	5.93	6	0.99	1.00	1.00	0.06	2283.48	2	0.78
K3_AS_2	M1		10.74**	2	5.37	0.98	0.85	0.01	2144.78		
	M2	M1	12.27	4	3.07	0.98	0.97	0.04	2138.53	2	2.05
	M3	M2	13.56**	6	2.26	0.98	0.92	0.04	2141.94	2	1.06
K1_AS_2	M1		1.27	3	0.42	1.00	1.02	0.01	1270.55		
	M2	M1	2.15	4	0.54	1.00	1.01	0.02	1268.76	2	0.92
	M3	M2	8.21	6	1.37	1.00	0.99	0.03	1271.64	2	4.83
	M1		3.63	3	1.21	1.00	1.00	0.01	1891.95		
	M2	M1	4.40	5	0.88	1.00	1.00	0.03	1889.17	2	0.57
	M3	M2	4.59	7	0.66	1.00	1.01	0.03	1885.52	2	0.14
K3_AS_2	M1		2.46	3	0.82	1.00	1.01	0.01	1749.21		
	M2	M1	6.68	5	1.34	1.00	0.99	0.06	1749.27	2	4.30
	M3	M2	10.76	7	1.54	0.99	0.99	0.05	1750.21	2	3.18

Anmerkungen. N(K1_AS_2) = 170; N(K2_AS_2) = 215; N(K3_AS_2) = 190; SE = Schulerfolg; Ma = Mathematik; De = Deutsch; CFI = Comparative Fit Index; TLI = Tucker-Lewis-Index; SRMR = Standardized Root Mean Square Residual; AIC = Akaike Information Criterion; Δdf = Differenz der Freiheitsgrade genesteter Modelle; $\Delta\chi^2(\text{TRd})$ = Satorra-Bentler Chi-Quadrat Differenztest; M1 = Bestfittendes Modell der Messmodellprüfung; M2 = Modell mit gleichgesetzten Faktorladungen; M3 = Modell mit gleichgesetzten Intercepts der Indikatoren; M4 = M2 mit freigesetzter Ladung Jahremnote 2004/05; Die letztlich ausgewählten Modelle der weiteren Analysen sind fett markiert.

* $p < 0.05$. ** $p < 0.01$.

Tabelle B.6
Prüfung der Messinvarianz Zeit – FSK Mathematik und Deutsch

	Modell	Vergleich	χ^2	df	χ^2/df	CFI	TLI	SRMR	AIC	Δdf	$\Delta\chi^2(TRd)$	
FSK_Ma	K1_AS_2	M1	9.21	15	0.61	1.00	1.04	0.03	3052.77			
	M2	M1	11.62	19	0.61	1.00	1.04	0.05	3048.03	4	1.79	
	M3	M2	21.28	23	0.93	1.00	1.00	0.06	3048.36	4	8.48	
	K2_AS_2	M1		27.39*	15	1.83	0.98	0.94	0.05	3773.37		
		M2	M1	43.56**	19	2.29	0.95	0.91	0.10	3782.27	4	15.51**
		M4	M1	29.17*	17	1.72	0.98	0.98	0.05	3772.09	2	1.38
	K3_AS_2	M1		20.28	15	1.35	0.99	0.97	0.05	3338.82		
		M2	M1	25.08	19	1.32	0.99	0.98	0.06	3335.96	4	4.44
		M3	M2	33.02	23	1.44	0.98	0.97	0.07	3336.21	4	8.00
FSK_De	K1_AS_2	M1	13.29	15	0.89	1.00	1.02	0.04	3015.69			
	M2	M1	15.42	19	0.81	1.00	1.03	0.06	3009.93	4	1.95	
	M3	M2	31.67	23	1.38	0.96	0.94	0.10	3021.34	4	13.01**	
	K2_AS_2	M1		26.76*	15	1.78	0.97	0.94	0.05	4004.42		
		M2	M1	32.64*	19	1.72	0.97	0.94	0.07	4005.46	4	3.45
		M3	M2	43.96**	23	1.91	0.95	0.93	0.07	4009.25	4	11.22*
	K3_AS_2	M1		45.59**	15	3.04	0.94	0.86	0.05	3396.11		
		M2	M1	50.23**	19	2.64	0.94	0.88	0.07	3396.64	4	3.55
		M3	M2	61.90**	23	2.69	0.92	0.88	0.09	3401.99	4	10.08*

Anmerkungen. N(K1_AS_2) = 170; N(K2_AS_2) = 215; N(K3_AS_2) = 190; FSK = Fähigkeitsselbstkonzept; Ma = Mathematik; De = Deutsch; CFI = Comparative Fit Index; TLI = Tucker-Lewis-Index; SRMR = Standardized Root Mean Square Residual; AIC = Akaike Information Criterion; Δdf = Differenz der Freiheitsgrade genesteter Modelle; $\Delta\chi^2(TRd)$ = Satorra-Bentler Chi-Quadrat Differenztest; M1 = Best-fittendes Modell der Messmodellprüfung; M2 = Modell mit gleichgesetzten Faktorladungen; M3 = Modell mit gleichgesetzten Intercepts der Indikatoren; M4 = M2 mit freigesetzter Ladung von Item 2 (partielle Messinvarianz).

* $p < 0.05$.

** $p < 0.01$.

Tabelle B.7
Prüfung der Messinvarianz Zeit – IM Mathematik und Deutsch

	Modell	Vergleich	χ^2	df	χ^2/df	CFI	TLI	SRMR	AIC	Δdf	$\Delta\chi^2(TRd)$	
IM_Ma	K1_AS_2	M1	7.78	18	0.43	1.00	1.08	0.03	3478.49			
	M2	M1	11.72	22	0.53	1.00	1.07	0.04	3468.61	4	3.60	
	M3	M2	26.93	26	1.04	0.99	1.00	0.07	3475.66	4	16.38**	
	K2_AS_2	M1		12.69	18	0.71	1.00	1.00	0.03	4402.56		
		M2	M1	16.18	22	0.74	1.00	1.02	0.05	4399.29	4	2.53
		M3	M2	27.27	26	1.05	1.00	1.00	0.07	4404.28	4	10.52*
	K3_AS_2	M1		28.92*	18	1.61	0.98	0.95	0.06	3901.59		
		M2	M1	35.28*	22	1.60	0.97	0.95	0.07	3900.97	4	5.48
		M3	M2	53.67**	26	2.06	0.94	0.92	0.09	3913.05	4	18.51**
IM_De	K1_AS_2	M1	23.66**	10	2.37	0.94	0.83	0.06	3118.64			
	M2	M1	24.83*	13	1.91	0.94	0.89	0.07	3116.67	3	0.81	
	M3	M2	49.68**	16	3.11	0.85	0.75	0.10	3134.68	3	23.74**	
	K2_AS_2	M1		17.70	10	1.77	0.98	0.94	0.05	4127.37		
		M2	M1	31.22**	13	2.40	0.95	0.90	0.07	4135.01	3	14.23**
		M4	M1	23.17*	12	1.93	0.97	0.93	0.07	4130.32	2	4.33
	K3_AS_2	M1		8.32	10	0.83	1.00	1.01	0.03	3518.70		
		M2	M1	10.72	13	0.82	1.00	1.01	0.05	3515.22	3	2.02
		M3	M2	20.91	16	1.31	0.99	0.98	0.06	3520.57	3	9.22*

Anmerkungen. N(K1_AS_2) = 170; N(K2_AS_2) = 215; N(K3_AS_2) = 190; IM = Intrinsische Motivation; Ma = Mathematik; De = Deutsch; CFI = Comparative Fit Index; TLI = Tucker-Lewis-Index; SRMR = Standardized Root Mean Square Residual; AIC = Akaike Information Criterion; Δdf = Differenz der Freiheitsgrade generierter Modelle; $\Delta\chi^2(TRd)$ = Satorra-Bentler Chi-Quadrat Differenztest; M1 = Best-fittendes Modell der Messmodellprüfung; M2 = Modell mit gleichgesetzten Faktorladungen; M3 = Modell mit gleichgesetzten Intercepts der Indikatoren; M4 = M2 mit freigesetzter Ladung von Item 2 (partielle Messinvarianz).

* $p < 0.05$.

** $p < 0.01$.

Tabelle B.8
Manifeste Interkorrelation zwischen Motivation und Schulnoten

	Schulerfolg												
	K_AS_2			K1_AS_2			K2_AS_2			K3_AS_2			
	Z1	Z2	Z3	Z1	Z2	Z3	Z1	Z2	Z3	Z1	Z2	Z3	
FSK_Ma	Z1	0.57	0.31	0.30	0.54	0.27	0.21*	0.55	0.28	0.38	0.64	0.33	0.27
	Z2	0.26	0.40	0.28	0.23 ^{n.s.}	0.37	0.25*	0.21	0.37	0.27	0.29	0.40	0.31
	Z3	0.32	0.36	0.38	0.20 ^{n.s.}	-0.03 ^{n.s.}	0.25	0.29	0.31	0.46	0.35	0.55	0.37
FSK_De	Z1	0.61	0.47	0.40	0.48	0.45	0.32	0.60	0.45	0.45	0.69	0.45	0.42
	Z2	0.35	0.53	0.36	0.29*	0.58	0.28*	0.34	0.45	0.45	0.44	0.51	0.31
	Z3	0.33	0.41	0.36	-0.02 ^{n.s.}	0.33	0.32	0.43	0.40 ^{n.s.}	0.42	0.39	0.47	0.30
IM_Ma	Z1	0.40	0.25	0.24	0.33*	0.15 ^{n.s.}	0.15 ^{n.s.}	0.36	0.20*	0.26	0.46	0.34	0.29
	Z2	0.16	0.31	0.19	0.09 ^{n.s.}	0.24	0.20*	0.16*	0.29	0.15 ^{n.s.}	0.14 ^{n.s.}	0.30	0.19 ^{n.s.}
	Z3	0.15	0.20	0.27	0.12 ^{n.s.}	0.00 ^{n.s.}	0.31*	0.01 ^{n.s.}	0.09 ^{n.s.}	0.20*	0.25	0.36	0.30
IM_De	Z1	0.41	0.30	0.24	0.31	0.30	0.22	0.39	0.18*	0.19*	0.47	0.34	0.31
	Z2	0.27	0.39	0.21	0.14 ^{n.s.}	0.42	0.21*	0.31	0.27	0.17 ^{n.s.}	0.35	0.39	0.22
	Z3	0.16	0.24	0.20	-0.15 ^{n.s.}	0.20*	0.18^{n.s.}	0.14 ^{n.s.}	0.12 ^{n.s.}	0.20*	0.28	0.31	0.21*

Anmerkungen. N(K_AS_2) = 318-442; N(K1_AS_2) = 38-128; N(K2_AS_2) = 135-200; N(K3_AS_2) = 117-183; FSK = Fähigkeitsselbstkonzept; IM = Intrinsische Motivation; Ma = Mathematik; De = Deutsch; Alle Korrelationen signifikant mit $p < 0.01$, es sei denn explizit anders gekennzeichnet.

* $p < 0.05$. ^{n.s.} nicht signifikant.

Tabelle B.9

Full-Forward Modell Fähigkeitsselbstkonzept Mathematik und Deutsch mit Kontrolle von g

		χ^2	df	CFI	TLI	SRMR	RMSEA	p	AIC
FSK_Ma	M1	174.59**	119	0.99	0.98	0.06	0.03	1.00	23 746.28
	M1a	144.61	119	0.98	0.97	0.06	0.04	0.88	6813.78
	M1b	147.00*	119	0.98	0.97	0.07	0.03	0.95	8834.60
	M1c	183.40**	119	0.95	0.93	0.09	0.05	0.34	7938.30

Anmerkungen. FSK = Fähigkeitsselbstkonzept; Ma = Mathematik; De = Deutsch; CFI = Comparative Fit Index; TLI = Tucker-Lewis-Index; SRMR = Standardized Root Mean Square Residual; RMSEA = Root-Mean-Square-Error of Approximation; p = $P(RMSEA < 0.05)$; AIC = Akaike Information Criterion; M1 = Full Forward Modell für K_AS_2; M1a = Full Forward Modell für K1_AS_2; M1b = Full Forward Modell für K2_AS_2; M1c = Full Forward Modell für K3_AS_2.

* $p < 0.05$. ** $p < 0.01$.

C Materialanhang

Erhebung der Zygote im Elternfragebogen

(Deutsche Übersetzung, adaptiert nach Price et al. (2000a))

Sind die Zwillinge eineiig oder zweieiig?

Wenn die Zwillinge verschiedenen Geschlechts sind, gehen Sie bitte weiter zu Seite 21

Wir sind daran interessiert, ob Sie glauben, dass die Zwillinge eineiig oder zweieiig sind. Deshalb bitten wir, Sie die folgenden Fragen zu beantworten.

Äußerliche Unterschiede

- 1) **Gibt es Unterschiede bei der farblichen Schattierung der Haare der Zwillinge?**
 ein deutlicher Unterschied ein leichter Unterschied nein
- 2) **Gibt es Unterschiede in der Beschaffenheit der Haare der Zwillinge (z.B. fein oder dick, glatt oder gelockt)?**
 ein deutlicher Unterschied ein leichter Unterschied nein
- 3) **Gibt es Unterschiede in den Augenfarben der Zwillinge?**
 ein deutlicher Unterschied ein leichter Unterschied nein
- 4) **Gibt es Unterschiede in der Form der Ohrläppchen der Zwillinge?**
 ein deutlicher Unterschied ein leichter Unterschied nein
- 5) **Haben die Zwillinge in etwa zur gleichen Zeit Zähne bekommen?**
 Die entsprechenden Zähne auf derselben Seite sind bei beiden nur um wenige Tage versetzt gekommen.
 Die entsprechenden Zähne sind bei beiden nur um wenige Tage versetzt gekommen, allerdings auf entgegengesetzten Seiten.
 Verschiedene Zähne sind nur um wenige Tage versetzt gekommen.
 Die ersten Zähne der Zwillinge kamen deutlicher zeitversetzt.
- 6) **Kennen Sie die Blutgruppen (A, B, AB, 0) und Rhesusfaktoren (Rh) Ihrer Zwillinge?**
 ja nein

Wenn ja, kreuzen Sie bitte die jeweilige Blutgruppe und den jeweiligen Rhesusfaktor des Zwillinges an!

Älterer	<input type="checkbox"/> A	<input type="checkbox"/> B	<input type="checkbox"/> AB	<input type="checkbox"/> 0	<input type="checkbox"/> Rh+	<input type="checkbox"/> Rh-
Jüngerer	<input type="checkbox"/> A	<input type="checkbox"/> B	<input type="checkbox"/> AB	<input type="checkbox"/> 0	<input type="checkbox"/> Rh+	<input type="checkbox"/> Rh-

7) Gibt es Unterschiede zwischen den Zwillingen, die durch Probleme bei der Geburt, durch einen Unfall oder Krankheit verursacht sind?

' ja ' nein ' weiß nicht ' es gibt keine Unterschiede

Wenn ja, bitte beschreiben Sie näher:

.....

8) Die Ähnlichkeit der Zwillinge mit zunehmendem Alter:

' blieb gleich ' wurde weniger ' nahm zu

9) Wurde Ihnen jemals von medizinischem Personal (z.B. Hausarzt, Krankenschwester, Facharzt) die Eiigkeit Ihrer Zwillinge mitgeteilt?

' ja, eineiig ' ja, zweieiig ' nein

Wenn „ja“, wie kam diese Annahme zustande:

10) Was glauben Sie persönlich? Sind Ihre Zwillinge eineiig oder zweieiig?

' eineiig ' zweieiig

Wie kommen Sie zu dieser Annahme?

.....

11) Wenn Sie sich frühe Fotos Ihrer Zwillinge im Alter von 2-4 Jahren anschauen, können Sie die beiden auseinanderhalten, ohne Merkmale der Kleidung oder andere Hinweise heranzuziehen?

' ja, leicht ' ja, aber manchmal ist es schwierig ' nein, ich verwechsle sie sehr oft auf Fotos

12) Haben folgende Personen die Zwillinge im frühen Kindesalter verwechselt?

anderer Elternteil der Zwillinge	' ja, oft	' ja, manchmal	' selten/nie	' es gibt keinen anderen Elternteil
----------------------------------	-----------	----------------	--------------	-------------------------------------

ältere/s Geschwister	' ja, oft	' ja, manchmal	' selten/nie	' es gibt kein/e älteres/n Geschwister
----------------------	-----------	----------------	--------------	--

andere Verwandte	' ja, oft	' ja, manchmal	' selten/nie	' es gibt keine anderen Verwandten
------------------	-----------	----------------	--------------	------------------------------------

Babysitter/ Tagesmutter	' ja, oft	' ja, manchmal	' selten/nie	' es gibt keine/n Babysitter/ Tagesmutter
-------------------------	-----------	----------------	--------------	---

nahe Freunde der Eltern	' ja, oft	' ja, manchmal	' selten/nie	
-------------------------	-----------	----------------	--------------	--

entferntere Freunde der Eltern	' ja, oft	' ja, manchmal	' selten/nie	
--------------------------------	-----------	----------------	--------------	--

Menschen, die die Zwillinge zum ersten Mal sehen	' ja, oft	' ja, manchmal	' selten/nie	
--	-----------	----------------	--------------	--

13) Wurden die Zwillinge verwechselt, wenn sie zusammen waren?

' ja, oft ' ja, manchmal ' nein, so gut wie nie ' sie wurden nicht verwechselt

14) Würden Sie sagen, dass Ihre Zwillinge im frühen Kindesalter:

- ' sich äußerlich glichen, wie „ein Ei dem anderen“
- ' sich äußerlich so ähnlich waren, wie Geschwister es eben sind
- ' sich äußerlich überhaupt nicht ähnlich waren

Erhebung der allgemeinen kognitiven Fähigkeiten

Informationen zur Erhebung der allgemeinen kognitiven Fähigkeiten zu den einzelnen Messzeitpunkten können von der Autorin angefordert werden.

Erfassung des FSK und der IM

(Adaptation nach Eccles und Wigfield (2002), Wigfield und Eccles (2000))

Fachspezifisches FSK

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!

Wie gut bist du...	<i>gar nicht gut</i>	<i>nicht so gut</i>	<i>weder noch</i>	<i>ziemlich gut</i>	<i>sehr gut</i>
1. ... im Lesen?	<input type="radio"/>				
2. ... im Schreiben (z.B. Aufsätze)?	<input type="radio"/>				
3. ... in Rechtschreibung?	<input type="radio"/>				
4. ... im Lösen von Zahlenrätseln oder Textaufgaben?	<input type="radio"/>				
5. ... im Kopfrechnen?	<input type="radio"/>				

Fachspezifische IM

Auf den folgenden Seiten wirst du wieder Fragen über verschiedene Dinge finden, die mit dir und der Schule zusammenhängen.

Zunächst würden wir gerne wissen, wie gerne du gewisse Dinge hast. Versuche dazu ganz allgemein zu überlegen, wie sehr du die folgenden Dinge **magst**.

Wie sehr magst du...	<i>mag es überhaupt nicht</i>	<i>mag es nicht</i>	<i>weder noch</i>	<i>mag es</i>	<i>mag es sehr</i>
1. ...Lesen?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. ...Schreiben (z.B. Aufsätze)?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. ...Rechtschreibung?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. ...Zahlenrätsel oder Textaufgaben lösen?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. ...Kopfrechnen?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. ...Multiplizieren und Dividieren?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Erfassung der Schulnoten

Bitte tragen Sie hier die Noten der Zwillinge im Halbjahreszeugnis 05/06 ein:
 Falls Ihre Kinder in dem entsprechenden Fach nicht unterrichtet wurden, kreuzen Sie bitte „nicht erteilt“ an.

		1	2	3	4	5	6	Nicht erteilt
Mathematik	älterer Zwilling	<input type="radio"/>						
	jüngerer Zwilling	<input type="radio"/>						
Deutsch	älterer Zwilling	<input type="radio"/>						
	jüngerer Zwilling	<input type="radio"/>						
Deutsch – Rechtschreibung	älterer Zwilling	<input type="radio"/>						
	jüngerer Zwilling	<input type="radio"/>						
Deutsch – Aufsätze	älterer Zwilling	<input type="radio"/>						
	jüngerer Zwilling	<input type="radio"/>						
Deutsch – Lesen	älterer Zwilling	<input type="radio"/>						
	jüngerer Zwilling	<input type="radio"/>						
Sachkunde	älterer Zwilling	<input type="radio"/>						
	jüngerer Zwilling	<input type="radio"/>						
Erdkunde	älterer Zwilling	<input type="radio"/>						
	jüngerer Zwilling	<input type="radio"/>						
Biologie	älterer Zwilling	<input type="radio"/>						
	jüngerer Zwilling	<input type="radio"/>						
Geschichte	älterer Zwilling	<input type="radio"/>						
	jüngerer Zwilling	<input type="radio"/>						

Literatur

- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*(3), 317–332. doi:10.1007/BF02294359
- Alarcón, M., Knopik, V. S. & DeFries, J. C. (2000). Covariation of mathematics achievement and general cognitive ability in twins. *Journal of School Psychology*, *38*(1), 63–77. doi:10.1016/S0022-4405(99)00037-0
- Allison, P. D. (2002). *Missing data*. Sage University papers series on quantitative applications in the social sciences, series no. 07-136. Thousand Oaks, CA: Sage. doi:10.1348/000711002159653
- 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*(3), 417–426. doi:10.1348/026151005X26804
- Alspaugh, J. W. (1998). Achievement loss associated with the transition to middle school and high school. *Journal of Educational Research*, *92*(1), 20–25. doi:10.1080/00220679809597572
- Amelang, M., Bartussek, D., Stemmler, G. & Hagemann, D. (2006). *Differentielle Psychologie und Persönlichkeitsforschung* (6. Aufl.). Stuttgart: Kohlhammer.
- Amelang, M. & Zielinski, W. (2002). *Psychologische Diagnostik und Intervention* (2., korrigierte, aktualisierte und überarbeitete Auflage). Berlin: Springer.
- Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A. (2001). *Intelligenz-Struktur-Test 2000 R*. Göttingen: Hogrefe. doi:10.1026//0049-8637.32.3.166
- Arens, A. K., Yeung, A. S., Craven, R. G. & Hasselhorn, M. (2011). The twofold multidimensionality of academic self-concept: Domain specificity and separation between competence and affect components. *Journal of Educational Psychology*, *103*(4), 970–981. doi:10.1037/a0025047
- Asbury, K., Wachs, T. D. & Plomin, R. (2005). Environmental moderators of genetic influence on verbal and nonverbal abilities in early childhood. *Intelligence*, *33*(6), 643–661. doi:10.1016/j.intell.2005.03.008
- Asendorpf, J. B. (2004). *Psychologie der Persönlichkeit* (3. Aufl.). Berlin: Springer.
- Atkinson, J. W. (1957). Motivational determinants of risk taking behavior. *Psychological Review*, *64*, 359–372. doi:10.1037/h0043445

- Ball, J., Lohaus, A. & Miebach, C. (2006). Psychische Anpassung und schulische Leistungen beim Wechsel von der Grundschule in die weiterführende Schule. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *38*(3), 101–109. doi:10.1026/0049-8637.38.3.101
- Baltes, P. B., Reese, H. W. & Lipsitt, L. P. (1980). Life-span developmental psychology. *Annual review of psychology*, *31*(1), 65–110. doi:10.1146/annurev.ps.31.020180.000433
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Barron, K. E. & Harackiewicz, J. M. (2001). Achievement goals and optimal motivation: Testing multiple goal models. *Journal of Personality and Social Psychology*, *80*(5), 706–722. doi:10.1037/0022-3514.80.5.706
- Bartels, M., Rietveld, M. J. H., Van Baal, G. C. M. & Boomsma, D. I. (2002a). Genetic and environmental influences on the development of intelligence. *Behavior Genetics*, *32*(4), 237–249. doi:10.1023/A:1019772628912
- Bartels, M., Rietveld, M. J. H., Van Baal, G. C. M. & Boomsma, D. I. (2002b). Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. *Twin Research*, *5*(6), 544–553. doi:10.1375/136905202762342017
- Baumert, J., Becker, M., Neumann, M. & Nikolova, R. (2009). Frühübergang in ein grundständiges Gymnasium – Übergang in ein privilegiertes Entwicklungsmilieu? *Zeitschrift für Erziehungswissenschaft*, *12*(2), 189–215.
- 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. 317–326). Kiel: IPN.
- Becker, M., Lüdtke, O., Trautwein, U. & Baumert, J. (2006). Leistungszuwachs in Mathematik. *Zeitschrift für Pädagogische Psychologie*, *20*(4), 233–242. doi:10.1024/1010-0652.20.4.233
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. doi:10.1037/0033-2909.107.2.238
- Bergen, S., Gardner, C. & Kendler, K. (2007). Age-related changes in heritability of behavioral phenotypes over adolescence and young adulthood: a meta-analysis. *Twin Research and Human Genetics*, *10*(3), 423–433. doi:10.1375/twin.10.3.423
- Binet, A. & Simon, T. (1904a). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'année psychologique*, *11*, 191–244. doi:10.3406/psy.1904.3675
- Binet, A. & Simon, T. (1904b). Sur la nécessité d'établir un diagnostic scientifique des états inférieurs de l'intelligence. *L'année psychologique*, *11*, 163–190. doi:10.3406/psy.1904.3674

- Binet, A. & Simon, T. (1905). Applications des méthodes nouvelles au diagnostic du niveau intellectuel chez des enfants normaux et anormaux d'hospice et d'école primaire. *L'année psychologique*, *11*, 245–336. doi:10.3406/psy.1904.3676
- Bishop, E. G., Cherny, S. S., Corley, R., Plomin, R., DeFries, J. C. & Hewitt, J. K. (2003). Development genetic analysis of general cognitive ability from 1 to 12 years in a sample of adoptees, biological siblings, and twins. *Intelligence*, *31*(1), 31–49. doi:10.1016/S0160-2896(02)00112-5
- Björklund, A., Jäntti, M. & Solon, G. (2005). Influences of nature and nurture on earnings variation: A report on a study of various sibling types in Sweden. In S. Bowles, H. Gintis & M. Osborne (Hrsg.), *Unequal chances: family background and economic success* (S. 145–164). New York: Russell Sage Foundation.
- Bloom, B. S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.
- Blyth, D. A., Simmons, R. G. & Carlton-Ford, S. (1983). The adjustment of early adolescents to school transitions. *Journal of Early Adolescence*, *3*, 105–120. doi:10.1177/027243168331008
- Böhnel, E. (1996). Die Frage der Prognostizierbarkeit von Schulerfolg in der Sekundarstufe I aufgrund der Benotung in der Primarstufe. *Unterrichtswissenschaft*, *24*, 343–360.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Boomsma, D. I. & Molenaar, P. C. M. (1987). The genetic analysis of repeated measures I: Simplex models. *Behavior Genetics*, *17*(2), 111–123. doi:10.1007/BF01065991
- Borkenau, P. (1993). *Anlage und Umwelt. Eine Einführung in die Verhaltensgenetik*. Göttingen: Hogrefe.
- Borkenau, P., Riemann, R., Angleitner, A. & Spinath, F. M. (2002). Similarity of childhood experiences and personality resemblance in monozygotic and dizygotic twins: A test of the equal environments assumption. *Personality and Individual Differences*, *33*(2), 261–269. doi:10.1016/S0191-8869(01)00150-7
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler* (5., vollständig überarbeitete und aktualisierte Auflage). Berlin: Springer.
- Bouchard, T. J. & Propping, P. (1993). *Twins as a tool of behavioral genetics*. Chichester: John Wiley & Sons.
- Bouchard, T. & McGue, M. (1981). Familial studies of intelligence: A review. *Science*, *212*(4498), 1055–1059. doi:10.1126/science.7195071
- Bouffard, T., Marcoux, M.-F., Vezeau, C. & Bordeleau, L. (2003). Changes in self-perceptions of competence and intrinsic motivation among elementary school-children. *British Journal of Educational Psychology*, *73*(2), 171–186. doi:10.1348/00070990360626921
- Branden, N. (1994). *Six pillars of self-esteem*. New York: Bantam.

- Brocke, B. & Beauducel, A. (2001). Intelligenz als Konstrukt. In E. Stern & J. Guthke (Hrsg.), *Perspektiven der Intelligenzforschung* (S. 13–42). Lengerich: Pabst Science Publishers.
- Brocke, B., Spinath, F. M. & Strobel, A. (2004). Verhaltensgenetische Ansätze und Theorien der Persönlichkeitsforschung. In K. Pawlik (Hrsg.), *Theorien und Anwendungsfelder der Differentiellen Psychologie* (S. 431–486). Band 5, Serie VIII der Enzyklopädie der Psychologie. Göttingen: Hogrefe.
- Brunner, M., Keller, U., Dierendonck, C., Reichert, M., Ugen, S., Fischbach, A. & Martin, R. (2010). The structure of academic self-concepts revisited: The nested Marsh/Shavelson model. *Journal of Educational Psychology*, *102*(4), 964–981. doi:10.1037/a001964
- Brunstein, J. & Heckhausen, H. (2006). Leistungsmotivation. In J. Heckhausen & H. Heckhausen (Hrsg.), *Motivation und Handeln* (S. 143–191). Berlin: Springer. doi:10.1007/3-540-29975-0_6
- Büchner, P. & Koch, K. (2001). *Der Übergang aus Kinder- und Elternsicht*. Stuttgart: Leske & Budrich.
- Buff, A. (1991). Schulische Selektion und Selbstkonzeptentwicklung. In H. Pekrun R. und Fend (Hrsg.), *Schule und Persönlichkeitsentwicklung: Ein Resumée der Längsschnittforschung* (S. 100–114). Stuttgart: Enke.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion* (2. aktualisierte Auflage). München: Pearson Studium.
- Bühner, M. & Ziegler, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler*. München: Pearson Studium.
- Byrne, B. M. (1996). Academic self-concept: Its structure, measurement, and relation to academic achievement. In B. A. Bracken (Hrsg.), *Handbook of self-concept: Developmental, social, and clinical considerations* (S. 287–316). New York: Wiley.
- Byrne, B. M. (2001). *Structural equation modeling with Amos: Basic concepts, applications, and programming*. Mahwah, NJ: Erlbaum.
- Byrne, B., Wadsworth, S., Corley, R., Samuelsson, S., Quain, P., DeFries, J. C., ... Olson, R. K. (2005). Longitudinal twin study of early literacy development: Preschool and kindergarten phases. *Scientific Studies of Reading*, *9*(3), 219–235. doi:10.1207/s1532799xssr0903_3
- Byrne, B. M. & Shavelson, R. J. (1987). Adolescent self-concept: Testing the assumption of equivalent structure across gender. *American Educational Research Journal*, *24*(3), 365–385. doi:10.3102/00028312024003365
- Calsyn, R. J. & Kenny, D. A. (1977). Self-concept of ability and perceived evaluation of others: Cause or effect of academic achievement. *Journal of Educational Psychology*, *69*(2), 136–145. doi:10.1037/0022-0663.69.2.136

- Carroll, J. B. (1993). *Human cognitive abilities. A survey of factor-analytic studies*. Cambridge: Cambridge University Press.
- Carroll, J. B. (1997). The three-stratum theory of cognitive abilities. In D. P. Flanagan, J. L. Genshaft & P. L. Harrison (Hrsg.), *Contemporary intellectual assessment: theories, tests, and issues* (S. 122–130). New York: Guilford Press.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1–22. doi:10.1037/h0046743
- Cattell, R. B., Barton, K. & Dielman, T. E. (1972). Prediction of school achievement from motivation, personality, and ability measures. *Psychological Reports*, *30*(1), 35–43. doi:10.2466/pr0.1972.30.1.35
- Cattell, R. B. & Butcher, H. J. (1968). *The prediction of achievement and creativity*. Indianapolis: Bobbs-Merill.
- Chamorro-Premuzic, T., Harlaar, N., Grewen, C. & 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*(4), 385–392. doi:10.1016/j.intell.2010.05.002
- Chen, F., Bollen, K. A., Paxton, P., Curran, P. J. & Kirby, J. B. (2001). Improper solutions in structural equation models. *Sociological Methods & Research*, *29*(4), 468–508. doi:10.1177/0049124101029004003
- Chiu, M. M. & Xihua, Z. (2008). Family and motivation effects on mathematics achievement: Analyses of students in 41 countries. *Learning and Instruction*, *18*(4), 321–336. doi:10.1016/j.learninstruc.2007.06.003
- Christ, O. & Schlüter, E. (2012). *Strukturgleichungsmodelle mit Mplus: Eine praktische Einführung*. München: Oldenbourg Wissenschaftsverlag.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Aufl.). Hillsdale: Lawrence Erlbaum Associates.
- Cohen, J., Cohen, P., West, S. & Aiken, L. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cole, M. S., Bedeian, A. G. & Feild, H. S. (2006). The measurement equivalence of web-based and paper-and-pencil measures of transformational leadership: A multinational test. *Organizational Research Methods*, *9*(3), 339–368. doi:10.1177/1094428106287434
- Cooley, W. W. & Lohnes, P. R. (1976). *Evaluation research in education*. New York: Irvington.
- Crockett, L. J., Petersen, A. C., Graber, J. A., Schulenberg, J. E. & Ebata, A. (1989). School transitions and adjustment during early adolescence. *The Journal of Early Adolescence*, *9*(3), 181–210. doi:10.1177/0272431689093002

- Davey, A., Shanahan, M. J. & Schafer, J. L. (2001). Correcting for selective nonresponse in the National Longitudinal Survey of Youth using multiple imputation. Zugriff am 18. August 2012, unter <http://www.jstor.org/stable/3069628>
- Davis, O. S. P., Arden, R. & Plomin, R. (2007). g in middle childhood: Moderate genetic and shared environmental influence using diverse measures of general cognitive ability at 7, 9 and 10 years in a large population sample of twins. *Intelligence*, *36*(1), 68–80. doi:10.1016/j.intell.2007.01.006
- De Geus, E. (2011). The association between depression and heart disease: The role of genetic factors. In A. Glassman, M. Maj & N. Sartorius (Hrsg.), *Depression and heart disease* (S. 57–98). Chichester: John Wiley und Sons, Ltd.
- Deary, I. J. (2001). *Intelligence. A very short introduction*. Oxford: Oxford University Press.
- Deary, I. J., Spinath, F. M. & Bates, T. C. (2006). Genetics of intelligence. *European Journal of Human Genetics*, *14*(6), 690–700. doi:10.1038/sj.ejhg.5201588
- Deary, I. J., Strand, S., Smith, P. & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*(1), 13–21. doi:10.1016/j.intell.2006.02.001
- Deary, I. J., Whalley, L. J., Lemmon, H., Crawford, J. & Starr, J. M. (2000). The stability of individual differences in mental ability from childhood to old age: Follow-up of the 1932 Scottish Mental Survey. *Intelligence*, *28*(1), 49–55. doi:10.1016/S0160-2896(99)00031-8
- Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J. & Fox, H. C. (2004). The impact of childhood intelligence on later life: Following up the Scottish mental surveys of 1932 and 1947. *Journal of Personality and Social Psychology*, *86*(1), 130–147. doi:10.1037/0022-3514.86.1.130
- Deci, E. L. & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.
- Deci, E. L. & Ryan, R. M. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik*, *39*(2), 223–238.
- Dickhäuser, O. & Galfe, E. (2004). Besser als..., schlechter als... *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *36*(1), 1–9. doi:10.1026/0049-8637.36.1.1
- Dickhäuser, O., Seidler, A. & Kölzer, M. (2005). Kein Mensch kann alles? Effekte dimensionaler Vergleiche auf das Fähigkeitsselbstkonzept. *Zeitschrift für pädagogische Psychologie*, *19*(1), 97–106. doi:10.1024/1010-0652.19.12.97
- Ditton, H. (2010a). Differentielle Leistungsentwicklung in der zweiten Hälfte der Grundschulzeit. *Zeitschrift für Grundschulforschung*, *3*, 83–98.
- Ditton, H. (2010b). Selektion und Exklusion im Bildungssystem. In G. Quenzel & K. Hurrelmann (Hrsg.), *Bildungsverlierer - Neue Ungleichheiten* (S. 53–72). Wiesbaden: VS Verlag für Sozialwissenschaften.

- Duchesne, S., Larose, S. & Guay, F. (2005). The transition from elementary to high school: The pivotal role of mother and child characteristics in explaining trajectories of academic functioning. *International Journal of Behavioral Development*, 29(5), 409–417. doi:10.1080/01650250500206067
- Duckworth, A., Quinn, P. & Tsukayama, E. (2011). What no child left behind leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades. *Journal of Educational Psychology*, Advance online publication. doi:10.1037/a0026280
- Dweck, C. S. (2002). The development of ability conceptions. In A. Wigfield & J. S. Eccles (Hrsg.), *Development of achievement motivation* (S. 57–88). San Diego, CA: Academic Press. doi:10.1016/B978-012750053-9/50005-X
- Dweck, C. S. & Legett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95, 256–273. doi:10.1037/0033-295X.95.2.256
- Eaves, L. J., Eysenck, H. & Martin, N. G. (1989). *Genes, culture, and personality: An empirical approach*. London: Academic Press.
- Eaves, L. J. & Gale, J. S. (1974). A method for analyzing the genetic basis of covariation. *Behavior Genetics*, 4(3), 253–267. doi:10.1007/BF01074158
- Eaves, L. J., Silberg, J. L. & Maes, H. H. (2005). Revisiting the children of twins: Can they be used to resolve the environmental effects of dyadic parental treatment on child behavior? *Twin Research and Human Genetics*, 8(4), 283–290. doi:10.1375/1832427054936736
- 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: Guilford Press.
- Eccles, 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. & Midgley, C. (1989). *Stage/environment fit: Developmentally appropriate classrooms for early adolescents* (R. Ames & C. Ames, Hrsg.). New York: Academic Press.
- Eccles, J. S. & Wigfield, A. (1995). In the mind of the achiever: The structure of adolescents academic achievement-related beliefs and self-perceptions. *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. doi:10.1146/annurev.psych.53.100901.135153
- Eccles, J. S., Wigfield, A. & Schiefele, U. (1998). Motivation to succeed. In W. Damon & N. Eisenberger (Hrsg.), *Handbook of child psychology* (5th, Bd. 3, S. 1017–1095). New York: John Wiley.

- Eccles, J. S., Wigfield, A., Harold, R. D. & Blumenfeld, P. (1993a). Age and gender differences in children's self- and task perceptions during elementary school. *Child Development*, 64(3), 830–847. doi:10.1111/j.1467-8624.1993.tb02946.x
- Eccles, J. S., Arberton, A., Buchanan, C. M., Janis, J., Flanagan, C., Harold, R., ... Reuman, D. (1993b). School and family effects on the ontogeny of children's interests, self-perceptions, and activity choices. In J. E. Jacobs (Hrsg.), *Nebraska Symposium on Motivation, 1992: Developmental perspectives on motivation* (S. 145–208). Lincoln, NE, US: University of Nebraska Press.
- Eder, R. A. & Mangelsdorf, S. C. (1997). The emotional basis of early personality development: Implications for the emergent self-concept. In R. Hogan, J. Johnson & S. Briggs (Hrsg.), *Handbook of personality psychology* (S. 209–240). San Diego, CA: Academic Press. doi:10.1016/B978-012134645-4/50010-X
- Elliot, A. J. & McGregor, H. A. (2001). A 2 x 2 achievement goal framework. *Journal of Personality and Social Psychology*, 80(3), 501–519. doi:10.1037/0022-3514.80.3.501
- Falconer, D. S. (1960). *Introduction to quantitative genetics*. New York: Ronald Press.
- Felner, R. D., Primavera, J. & Cauce, A. M. (1981). The impact of school transitions: A focus for preventive efforts. *American Journal of Community Psychology*, 9(4), 449–459. doi:10.1007/BF00918175
- Filipp, S.-H. (1995). *Kritische Lebensereignisse*. München: Psychologie Verlags Union. doi:10.1026/0049-8637.38.3.101
- Filipp, S.-H. (2006). Kommentar zum Schwerpunktthema: Entwicklung von Fähigkeitsselbstkonzepten. *Zeitschrift für Pädagogische Psychologie*, 20(1-2), 65–72. doi:10.1024/1010-0652.20.12.65
- Finkel, D. & McGue, M. (1998). Age differences in the nature and origin of individual differences in memory: A behavior genetic analysis. *International journal of aging & human development*, 47(3), 217–239. doi:10.2190/JEPX-G60A-6QK3-9DR3
- Finkel, D., Pedersen, N. L., McClearn, G. E., Plomin, R. & Berg, S. (1996). Cross-sequential analysis of genetic influences on cognitive ability in the Swedish Adoption/Twin Study of Aging. *Aging, Neuropsychology, and Cognition*, 3(1), 84–99. doi:10.1080/13825589608256614
- Finsterwald, M. (2006). *Motivation und Schulübergang. Bedingungsfaktoren eines erfolgreichen Übergangs auf weiterführende Schulen*. Berlin: Logos Verlag Berlin.
- 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. doi:10.1016/0883-0355(87)90035-8
- Freese, J. (2008). Genetics and the social science explanation of individual outcomes. *American Journal of Sociology*, 114(1), 1–35. doi:10.1086/592208

- Frey, M. C. & Detterman, D. K. (2004). Scholastic assessment or g: The relationship between the scholastic assessment test and general cognitive ability. *Psychological Science*, *15*(6), 373–378. doi:10.1111/j.0956-7976.2004.00687.x
- Frey, M. J. (2012). *Vorhersage von Schulerfolg durch Intelligenz und motivationale Prädiktoren bei Schülern der 5. Klasse*. Unveröffentlichte Diplomarbeit, Universität des Saarlandes. Saarbrücken.
- Funke, J. & Vaterrodt-Plünnecke, B. (2004). *Was ist Intelligenz*. München: Beck.
- Gagné, F. & St Père, F. (2001). When IQ is controlled, does motivation still predict achievement. *Intelligence*, *30*(1), 71–100. doi:10.1016/S0160-2896(01)00068-X
- Galton, F. (1869). *Hereditary genius*. London: Macmillan.
- Geiser, C. (2010). *Datenanalyse mit Mplus: Eine anwendungsorientierte Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gerrig, R. J. & Zimbardo, P. G. (2008). *Psychologie* (18. Aufl.). München: Pearson Studium.
- Goldsmith, H. H. (1991). A zygoty questionnaire for young twins: A research note. *Behavior Genetics*, *21*, 257–269. doi:10.1007/BF01065819
- Gose, A., Wooden, S. & Muller, D. (1980). The relative potential of self-concept and intelligence as predictors of achievement. *Journal of Psychology*, *104*(2), 279–287.
- Gottfredson, L. S. (1997a). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, *24*(1), 13–23.
- Gottfredson, L. S. (1997b). Why g matters: The complexity of everyday life. *Intelligence*, *24*(1), 79–132. doi:10.1016/S0160-2896(97)90014-3
- Gottfredson, L. S. (2002). Where and why g matters: Not a mystery. *Human Performance*, *15*(1/2), 25–46. doi:10.1080/08959285.2002.9668082
- Gottfried, A. E., Marcoulides, G. A., Gottfried, A. W., Oliver, P. H. & Guerin, D. W. (2007). Multivariate latent change modeling of developmental decline in academic intrinsic math motivation and achievement: Childhood through adolescence. *International Journal of Behavioral Development*, *31*(4), 317. doi:10.1177/0165025407077752
- Gottfried, A. E. (September 1990). Academic intrinsic motivation in young elementary school children. *Journal of Educational Psychology*, *82*(3), 525–538. doi:10.1037/0022-0663.82.3.525
- Gottfried, A. E., Fleming, J. S. & Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: A longitudinal study. *Journal of Educational Psychology*, *93*(1), 3–13. doi:10.1037/0022-0663.93.1.3
- Gottschling, J., Spengler, M., Spinath, B. & Spinath, F. M. (2012). The prediction of school achievement from a behavior genetic perspective: Results from the German twin study on Cognitive Ability, Self-Reported Motivation, and School

- Achievement (CoSMoS). *Personality and Individual Differences*, 53(4), 381–386. doi:10.1016/j.paid.2012.01.020
- Grant, M. D., Kremen, W. S., Jacobson, K. C., Franz, C., Xian, H., Eisen, S. A., ... Lyons, M. J. (2010). Does parental education have a moderating effect on the genetic and environmental influences of general cognitive ability in early adulthood. *Behavior Genetics*, 40(4), 438–446. doi:10.1007/s10519-010-9351-3
- 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. doi:10.1111/j.1467-9280.2009.02366.x
- 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). Waxman.
- Guay, F., Larose, S. & Boivin, M. (2004). Academic self-concept and educational attainment level: A ten-year longitudinal study. *Self and Identity*, 3(1), 53–68. doi:10.1080/13576500342000040
- 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. doi:10.1037/0022-0663.95.1.124
- 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: Macmillan Library Reference USA.
- Hacker, H. (1988). Übergänge fordern uns heraus. *Die Grundschule*, 10, 8–10. doi:10.1007/978-3-531-91095-6_22
- Hahn, E., Spinath, F. M., Siedler, T., Wagner, G. G., Schupp, J. & Kandler, C. (2012). The complexity of personality: Advantages of a genetically sensitive multi-group design. *Behavior genetics*, 42(2), 221–233. doi:10.1007/s10519-011-9493-y
- Hanscombe, K. B., Trzaskowski, M., Haworth, C. M. A., Davis, O. S. P., Dale, P. S. & Plomin, R. (2012). Socioeconomic status (SES) and children's intelligence: A large UK-representative twin study finds SES moderation of the environmental, not genetic, effect on IQ. *PLoS ONE*, 7(2), 1–16. doi:10.1371/journal.pone.0030320
- Hansford, B. C. & Hattie, J. A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research*, 52(1), 123–142. doi:10.3102/00346543052001123
- Harackiewicz, J. M., Barron, K. E., Tauer, J. M. & Elliot, A. J. (2002). Predicting success in college: A longitudinal study of achievement goals and ability measures as predictors of interest and performance from freshman year through graduation.

- Journal of Educational Psychology*, 94(3), 562–575. doi:10.1037/0022-0663.94.3.562
- Harden, K., Turkheimer, E. & Loehlin, J. (2007). Genotype by environment interaction in adolescents cognitive aptitude. *Behavior Genetics*, 37(2), 273–283. doi:10.1007/s10519-006-9113-4
- Harris, J. R. (1995). Where is the child's environment? A group socialization theory of development. *Psychological Review*, 102(3), 458–489. doi:10.1037/0033-295X.102.3.458
- Hart, S. A. & Petrill, S. A. (2009). The genetics and environments of reading: A behavioral genetic perspective. In Y. . Kim (Hrsg.), *Handbook of Behavior Genetics* (S. 113–123). Springer. doi:10.1007/978-0-387-76727-7_8
- Harter, S. (1981). A new self-report scale of intrinsic versus extrinsic orientation in the classroom: Motivational and informational components. *Developmental Psychology*, 17(3), 300–312. doi:10.1037/0012-1649.17.3.300
- Harter, S. (1985). *Manual for the self-perception profile for children*. Denver: University of Denver.
- Harter, S. (1998). The development of self-representations. In W. Damon & N. Eisenberger (Hrsg.), *Handbook of child psychology* (5. Aufl., Bd. 3. Social, emotional, and personality development, S. 553–617). Wiley.
- Harter, S. (1999). *The construction of the self: A developmental perspective*. New York: The Guilford Press.
- Hasselhorn, M. & Grube, D. (1997). Entwicklung der Intelligenz und des Denkens. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 15–26). Weinheim: Beltz.
- Hattie, J. A. (1992). *Self-concept*. Hillsdale, NJ: Erlbaum.
- Havighurst, R. J. (1972). *Developmental task and education* (3rd). New York: Davis McKay.
- Haworth, C. M. A., Wright, M. J., Luciano, M., Martin, N. G., de Geus, E. J. C., van Beijsterveldt, C. E. M., ... Thompson, L. A. (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular Psychiatry*, 15(11), 1112–1120. doi:10.1038/mp.2009.55
- Heckhausen, H. (1980). *Fähigkeit und Motivation in erwartungswidriger Schulleistung*. Göttingen: Verlag für Psychologie.
- Heckhausen, H. (2006). Entwicklungslinien der Motivationsforschung. In J. Heckhausen & H. Heckhausen (Hrsg.), *Motivation und Handeln* (S. 77–86). Berlin: Springer. doi:10.1007/3-540-29975-0_2
- Heckhausen, J. & Heckhausen, H. (2006). *Motivation und Handeln*. Berlin: Springer Verlag.

- Hegewald, K. (2009). *Elterliches Involvement als Prädiktor des Schulerfolgs: Besteht ein inkrementeller Beitrag über die allgemeine Intelligenz hinaus*. Unveröffentlichte Diplomarbeit, Universität des Saarlandes. Saarbrücken.
- Hellbrück, I. (2007). *Lern- und Leistungsziele und Motivationsdispositionen bei Grundschulkindern: Inkrementelle prädiktive Validität über allgemeine Intelligenz hinaus und verhaltensgenetische Aspekte zur Ätiologie - Daten einer Zwillingsstudie*. Saarbrücken: Unveröffentlichte Diplomarbeit.
- Heller, K. A. (1997). Individuelle Bedingungsfaktoren der Schulleistung: Literaturüberblick. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 183–201). Weinheim: Beltz.
- Heller, K. A. (2003). Das Gymnasium zwischen Tradition und modernen Bildungsansprüchen. *Zeitschrift für Pädagogik*, 49, 213–234.
- Heller, K. A. & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision (KFT 4-12+R)*. Weinheim: Beltz.
- Heller, K. A. & Ziegler, A. (1998). Motivationsförderung im Unterricht. *Psychologie in Erziehung und Unterricht*, 45, 161–167.
- Helmke, A. (1992). *Selbstvertrauen und schulische Leistung*. Göttingen: Hogrefe.
- Helmke, A. (1998). Vom Optimisten zum Realisten? Zur Entwicklung des Fähigkeits-selbstkonzeptes vom Kindergarten bis zur sechsten Klassenstufe. In F. E. Weinert (Hrsg.), *Entwicklung im Kindesalter* (S. 115–132). Weinheim: Beltz.
- Helmke, A. & Schrader, F.-W. (2007). Entwicklung akademischer Leistungen. In M. Hasselhorn & W. Schneider (Hrsg.), *Handbuch der Entwicklungspsychologie* (S. 289–298). Handbuch der Psychologie. Göttingen: Hogrefe.
- Helmke, A. (1993). Die Entwicklung der Lernfreude vom Kindergarten bis zur 5. Klassenstufe. *Zeitschrift für Pädagogische Psychologie*, 7(2-3), 77–86.
- Helmke, A. (1997). Individuelle Bedingungsfaktoren der Schulleistung: Ergebnisse aus dem SCHOLASTIK-Projekt. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 203–221). Weinheim: Beltz.
- Helmke, A. & Schrader, F.-W. (2006). Determinanten der Schulleistung. In D. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 83–94). Weinheim: Beltz.
- Helmke, A., Schrader, F.-W. & Lehneis-Klepper, G. (1991). Zur Rolle des Elternverhaltens für die Schulleistungsentwicklung ihrer Kinder. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 23(1), 1–22.
- Helmke, A. & van Aken, M. A. G. (1995). The causal ordering of academic achievement and self-concept of ability during elementary school: A longitudinal study. *Journal of Educational Psychology*, 87(4), 624–637. doi:10.1037/0022-0663.87.4.624
- Helmke, A. & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen. In A. Helmke & F. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule* (3, S. 71–176). Göttingen: Hogrefe.

- Hill, M. A. (1997). *SPSS missing value analysis 7.5*. Chicago, Ill.: Spss Inc.
- Holling, H., Preckel, F. & Vock, M. (2004). *Intelligenzdiagnostik*. Göttingen: Hogrefe.
- Holodynski, M. (2007). Entwicklung der Leistungsmotivation. In M. Hasselhorn & W. Schneider (Hrsg.), *Handbuch der Entwicklungspsychologie*. Göttingen: Hogrefe.
- Horswill, M. & Coster, M. (2001). User-controlled photographic animations, photograph-based questions, and questionnaires: three Internet-based instruments for measuring drivers' risk-taking behavior. *Behavior Research Methods, Instruments, & Computers: A Journal Of The Psychonomic Society, Inc.*, 33(1), 46–58. doi:10.3758/BF03195346
- Howe, M. J. A. (1987). Motivation, cognition and individual achievements. In E. De Corte, H. Lodewijks, R. Parmentier & P. Span (Hrsg.), *Learning and Instruction: European Research in an International Context* (Bd. 1, S. 133–146). Oxford: Pergamon.
- Hoyle, R. H. (1995). The structural equation modeling approach: Basic concepts and fundamental issues. In R. H. Hoyle (Hrsg.), *Structural equation modeling: Concepts, issues, and applications* (S. 16–36). Thousand Oaks, CA: Sage Publications.
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. doi:10.1080/10705519909540118
- Huber, C., Späni, M., Schmellentin, C. & Criblez, L. (2006). *Bildungsstandards in Deutschland, Österreich, England, Australien, Neuseeland und Südostasien. Literaturbericht zu Entwicklung, Implementierung und Gebrauch von Standards in nationalen Bildungssystemen*. Aarau: Fachhochschule Nordwestschweiz.
- Huguet, P., Dumas, F., Marsh, H., Régner, I., Wheeler, L., Suls, J., ... Nezlek, J. (2009). Clarifying the role of social comparison in the big-fish-little-pond effect (BFLPE): An integrative study. *Journal of Personality and Social Psychology*, 97(1), 156–170. doi:10.1037/a0015558
- 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, New Jersey: Erlbaum.
- Hur, Y. M., McGue, M. & Iacono, W. G. (1998). The structure of self-concept in female preadolescent twins: A behavioral genetic approach. *Journal of Personality and Social Psychology*, 74(4), 1069–1077. doi:10.1037/0022-3514.74.4.1069
- Husén, T. (1959). *Psychological twin research*. Stockholm: Almqvist und Wiksell.
- Iacono, W. G. & McGue, M. (2002). Minnesota twin family study. *Twin Research and Human Genetics*, 5(5), 482–487. doi:10.1375/136905202320906327
- Iervolino, A. C., Pike, A., Manke, B., Reiss, D., Hetherington, E. M. & Plomin, R. (2002). Genetic and environmental influences in adolescent peer socialization:

- Evidence from two genetically sensitive designs. *Child Development*, *73*(1), 162–174. doi:DOI:10.1111/1467-8624.00398
- Ingenkamp, K. (1971). *Die Fragwürdigkeit der Zensurengebung*. Weinheim: Beltz.
- Jackson, D. L. (2003). Revisiting sample size and number of parameter estimates: Some support for the N:q hypothesis. *Structural Equation Modeling*, *10*(1), 128–141. doi:10.1207/S15328007SEM1001_6
- Jacobs, J. E., Lanza, S., Osgood, W., Eccles, J. S. & Wigfield, A. (2002). Changes in childrens self-competence and Values: Gender and domain differences across grades one through twelve. *Child Development*, *73*(2), 509–527. doi:10.1111/1467-8624.00421
- Jaffee, S. R. & Price, T. S. (2007). Gene–environment correlations: a review of the evidence and implications for prevention of mental illness. *Molecular Psychiatry*, *12*(5), 432–442. doi:10.1038/sj.mp.4001950
- 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.
- Jang, K. L., McCrae, R. R., Angleitner, A., Riemann, R. & Livesley, W. J. (1998). Heritability of facet-level traits in a cross-cultural twin sample: Support for a hierarchical model of personality. *Journal of Personality and Social Psychology*, *74*(6), 1556–1565. doi:10.1037/0022-3514.74.6.1556
- Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement. *Harvard Educational Review*, *39*(1), 1–123.
- Jensen, A. R. (1993). Psychometric g and achievement. In B. R. Gifford (Hrsg.), *Policy perspectives on educational testing* (S. 117–227). Boston, Dordrecht, London: Kluwer Academic Publishers.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport: Praeger.
- Jensen, A. R. (2002). Psychometric g: Definition and substantiation. In R. J. Sternberg & E. Grigorenko (Hrsg.), *The general factor of Intelligence. How general is it* (S. 39–53). Mahwah, New Jersey: Erlbaum.
- Jimerson, S., Egeland, B. & Teo, A. (1999). A longitudinal study of achievement trajectories: Factors associated with change. *Journal of Educational Psychology*, *91*(1), 116–126. doi:10.1037/0022-0663.91.1.116
- Johnson, W., McGue, M. & Iacono, W. G. (2005). Disruptive behavior and school grades: Genetic and environmental relations in 11-year-olds. *Journal of educational psychology*, *97*(3), 391–405. doi:10.1037/0022-0663.97.3.391
- Johnson, W., McGue, M. & Iacono, W. G. (2006). Genetic and environmental influences on academic achievement trajectories during adolescence. *Developmental Psychology*, *42*(3), 514–531. doi:10.1037/0012-1649.42.3.514

- Johnson, W., Nijenhuis, J. & Bouchard, T. J. (2008). Still just 1 g: Consistent results from five test batteries. *Intelligence*, *36*(1), 81–95. doi:10.1016/j.intell.2007.06.001
- Johnson, W., Penke, L. & Spinath, F. M. (2011). Heritability in the era of molecular genetics: Some thoughts for understanding genetic influences on behavioural traits. *European Journal of Personality*, *25*(4), 254–266. doi:10.1002/per.836
- Johnson, W., McGue, M. & Iacono, W. G. (2007). How parents influence school grades: Hints from a sample of adoptive and biological families. *Learning and Individual Differences*, *17*(3), 201–219. doi:10.1016/j.lindif.2007.04.004
- Joinson, A. (1999). Social desirability, anonymity, and internet-based questionnaires. *Behavior Research Methods, Instruments, & Computers: A Journal Of The Psychonomic Society, Inc*, *31*(3), 433–438. doi:10.3758/BF03200723
- Karbach, J., Gottschling, J., Spengler, M., Hegewald, K. & Spinath, F. M. (2013). Parental involvement and general cognitive ability as predictors of domain-specific academic achievement in early adolescence. *Learning and Instruction*, *23*, 43–51. doi:http://dx.doi.org/10.1016/j.learninstruc.2012.09.004
- Keller, M. C. & Coventry, W. L. (2005). Quantifying and addressing parameter indeterminacy in the classical twin design. *Twin Research and Human Genetics*, *8*(3), 201–213. doi:10.1375/1832427054253068
- Keller, M. C., Medland, S. E., Duncan, L. E., Hatemi, P. K., Neale, M. C., Maes, H. H. & Eaves, L. J. (2009). Modeling extended twin family data I: Description of the Cascade model. *Twin Research and Human Genetics*, *12*(1), 8–18. doi:10.1375/twin.12.1.8
- Kendler, K. S., Gardner, C. O. & Prescott, C. A. (1998). A population-based twin study of self-esteem and gender. *Psychological Medicine*, *28*(6), 1403–1409. doi:http://dx.doi.org/
- Kenny, D. A. (1975). Cross-lagged panel correlation: A test for spuriousness. *Psychological Bulletin*, *82*(6), 887–903. doi:10.1037/0033-2909.82.6.887
- Kenny, D. A. (4. September 2011). Measuring Model Fit. Zugriff am 30. Juni 2012, unter <http://davidakenny.net/cm/fit.htm>
- Klauer, K. J. (2006). Intelligenz und Begabung. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 83–94). Weinheim: Beltz.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd). New York: The Guilford Press.
- Klingbeil, W. (2004). *Das Fähigkeitsselbstkonzept von Grundschulkindern: Eine geschlechtsspezifische Betrachtung im Kontext schulischen Leistungsverhaltens*. Bielefeld: Unveröffentlichte Diplomarbeit der Fachrichtung Psychologie.
- Klug, W. S., Cummings, M. R. & Spencer, C. A. (2007). *Genetik* (8. Auflage). München: Pearson Studium.

- Klump, K. L., Holly, A., Iacono, W. G., McGue, M. & Willson, L. E. (2000). Physical similarity and twin resemblance for eating attitudes and behaviors: A test of the equal environments assumption. *Behavior Genetics*, *30*(1), 51–58. doi:10.1023/A:1002038610763
- Koch, K. (2004). Von der Grundschule zur Sekundarstufe. In W. Helsper & J. Böhme (Hrsg.), *Handbuch der Schulforschung* (S. 549–565). Wiesbaden: VS Verlag für Sozialwissenschaften. doi:10.1007/978-3-531-91095-6_22
- Köller, O. (2004). *Konsequenzen von Leistungsgruppierungen*. Münster: Waxmann.
- Köller, O., Daniels, Z., Schnabel, K. U. & Baumert, J. (2000). Kurswahlen von Mädchen und Jungen im Fach Mathematik: Zur Rolle von fachspezifischem Selbstkonzept und Interesse. *Zeitschrift für Pädagogische Psychologie*, *14*(1), 26–37. doi:10.1024//1010-0652.14.1.26
- Köller, O., Baumert, J. & Schnabel, K. (2001). Does interest matter? The relationship between academic interest and achievement in mathematics. *Journal for Research in Mathematics Education*, *32*(5), 448–470. doi:10.2307/749801
- Kovas, Y., Haworth, C. M. A., Dale, P. S. & Plomin. (2007). The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monographs Of The Society For Research In Child Development*, *72*(3), 1–144. doi:10.1111/j.1540-5834.2007.00439.x
- Krapp, A. (1999). Intrinsische Lernmotivation und Interesse. Forschungsansätze und konzeptuelle Überlegungen. *Zeitschrift für Pädagogik*, *45*(3), 387–406.
- Krapp, A. (2005). Basic needs and the development of interest and intrinsic motivational orientations. *Learning and Instruction*, *15*(5), 381–395. doi:10.1016/j.learninstruc.2005.07.007
- Krapp, A. (2006). Interesse. In D. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 281–290). Weinheim: Beltz.
- Kubinger, K. D. & Wurst, E. (2000). *Adaptives Intelligenz Diagnostikum 2 (AID 2)*. Göttingen: Beltz Test GmbH.
- 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. doi:10.1037/0022-3514.86.1.148
- Lazarsfeld, P. F. (1972). Mutual relations over time of two attributes: A review and integration of various approaches. In M. Hammer, K. Salzinger & S. Sutton (Hrsg.), *Psychopathology* (S. 461–480). New York: Wiley.
- Lehmann, R. & Lenkeit, J. (2008). *ELEMENT: Erhebung zum Lese- und Mathematikverständnis in den Jahrgangsstufen 4 bis 6 in Berlin. Abschlussbericht über die Untersuchungen 2003, 2004 und 2005 an Berliner Grundschulen und grundständigen Gymnasien*. Humboldt Universität zu Berlin.

- Lens, W. (1987). Motivation, cognition and individual achievements. In E. De Corte, H. Lodewijks, R. Parmentier & P. Span (Hrsg.), *Learning and Instruction: European Research in an International Context* (Bd. 1, S. 181–189). Oxford: Pergamon.
- Leve, L. D., Harold, G. T., Ge, X., Neiderhiser, J. M. & Patterson, G. (2010). Refining intervention targets in family-based research. *Perspectives on Psychological Science*, 5(5), 516–526. doi:10.1177/1745691610383506
- Levpušček, M. P. & Zupančič, M. (2009). Math achievement in early adolescence. *The Journal of Early Adolescence*, 29(4), 541–570. doi:10.1177/0272431608324189
- Lewin, K. (1935). *A dynamic theory of personality: Selected papers*. New York: McGraw Hill.
- Lienert, G. A. (1989). *Testaufbau und Testanalyse*. München: Psychologie Verlags Union.
- Light, J. G., Defries, J. C. & Olson, R. K. (1998). Multivariate behavioral genetic analysis of achievement and cognitive measures in reading-disabled and control twin pairs. *Human Biology*, 70(2), 215–237.
- Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83(404), 1198–1202. doi:10.2307/2290157
- Little, T. D., Oettingen, G., Stetsenko, A. & Baltes, P. B. (1995). Children's action-control beliefs about school performance: How do American children compare with German and Russian children. *Journal of Personality and Social Psychology*, 69, 686–700. doi:10.1037/0022-3514.69.4.686
- Little, T. D., Preacher, K. J., Selig, J. P. & Card, N. A. (2007). New developments in latent variable panel analyses of longitudinal data. *International Journal of Behavioral Development*, 31(4), 357–365. doi:10.1177/0165025407077757
- Lloyd, J. & Barenblatt, L. (1984). Intrinsic intellectuality: Its relations to social class, intelligence, and achievement. *Journal of Personality and Social Psychology*, 46(3), 646–654. doi:10.1037/0022-3514.46.3.646
- Loehlin, J. C. (1996). The Cholesky approach: A cautionary note. *Behavior Genetics*, 26(1), 65–69. doi:10.1007/BF02361160
- Loehlin, J. C. & Nichols, J. (1976). *Heredity, environment and personality*. Austin: University of Texas Press.
- Lohaus, A., Vierhaus, M. & Maass, A. (2010). *Entwicklungspsychologie des Kindes- und Jugendalters*. Heidelberg: Springer.
- Lüdtke, O. & Köller, O. (2002). Individuelle Bezugsnormorientierung und soziale Vergleiche im Mathematikunterricht. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 34(3), 156–166. doi:10.1026//0049-8637.34.3.156
- Lüdtke, O., Köller, O., Artelt, C., Stanat, P. & Baumert, J. (2002). Eine Überprüfung von Modellen zur Genese akademischer Selbstkonzepte: Ergebnisse aus der PISA-

- Studie. *Zeitschrift für Pädagogische Psychologie*, 16(3/4), 151–164. doi:10.1024/1010-0652.16.34.151
- Luo, D., Petrill, S. A. & Thompson, L. A. (1994). An exploration of genetic g: Hierarchical factor analysis of cognitive data from the Western Reserve Twin Project. *Intelligence*, 18(3), 335–347. doi:10.1016/0160-2896(94)90033-7
- Luo, Y. L. L., Haworth, C. M. A. & Plomin, R. (2010). A novel approach to genetic and environmental analysis of cross-lagged associations over time: The cross-lagged relationship between self-perceived abilities and school achievement is mediated by genes as well as the environment. *Twin Research and Human Genetics*, 13(5), 426–436. doi:10.1375/twin.13.5.426
- Luo, Y. L. L., Kovas, Y., Haworth, C. M. A. & Plomin, R. (2011). The etiology of mathematical self-evaluation and mathematics achievement: Understanding the relationship using a cross-lagged twin study from ages 9 to 12. *Learning and Individual Differences*, 21(6), 710–718. doi:10.1016/j.lindif.2011.09.001
- Maaz, K., Hausen, C., McElvany, N. & Baumert, J. (2006). Stichwort: Übergänge im Bildungssystem. *Zeitschrift für Erziehungswissenschaft*, 9, 299–327. doi:10.1007/s11618-006-0053-9
- Magnusson, D. & Backtemann, G. (1978). Longitudinal stability of person characteristics: Intelligence and creativity. *Applied Psychological Measurement*, 2(4), 841–849. doi:10.1177/014662167800200402
- Makel, M. C., Lee, S. Y., Olszewki-Kubilius, P. & Putallaz, M. (2012). Changing the pond, not the fish: Following high-ability students across different educational environments. *Journal of Educational Psychology*, 104(3), 778–792. doi:10.1037/a0027558
- Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology*, 79(3), 280–295. doi:10.1037/0022-0663.79.3.280
- Marsh, H. W. & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, 1(2), 133–163. doi:10.1111/j.1745-6916.2006.00010.x. eprint: <http://pps.sagepub.com/content/1/2/133.full.pdf+html>
- Marsh, H. W., Debus, R. & Bornholt, L. (2005a). Validating young children's self-concept responses: Methodological ways and means to understand their responses. In D. M. Teti (Hrsg.), *Handbook of research methods in developmental science* (S. 138–160). Oxford, England: Blackwell. doi:10.1002/9780470756676.ch8
- Marsh, H. W. & Hau, K. T. (1996). Assessing goodness of fit: Is parsimony always desirable. *The Journal of Experimental Educational*, 64(4), 364–390.
- Marsh, H. W. & Hau, K. T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the inter-

- nal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology*, *96*(1), 56–67. doi:10.1037/0022-0663.96.1.56
- Marsh, H. W. & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology*, *81*(1), 59–77. doi:10.1348/000709910X503501
- Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O. & Baumert, J. (2006). Integration of multidimensional self-concept and core personality constructs: Construct validation and relations to well-being and achievement. *Journal of Personality*, *74*(2), 403–456. doi:10.1111/j.1467-6494.2005.00380.x
- Marsh, H. W., Trautwein, U., Lüdtke, O., Baumert, J. & Köller, O. (2007). The big-fish-little-pond effect: Persistent negative effects of selective high schools on self-concept after graduation. *American Educational Research Journal*, *44*(3), 631–669. doi:10.3102/0002831207306728
- Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, *23*(1), 129–149. doi:10.3102/00028312023001129
- Marsh, H. W. (1990). Causal ordering of academic self-concept and academic achievement: A multiwave, longitudinal panel analysis. *Journal of Educational Psychology*, *82*(4), 646–656. doi:10.1037/0022-0663.82.4.646
- Marsh, H. W., Byrne, B. M. & Shavelson, R. J. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology*, *80*(3), 366–380. doi:10.1037/0022-0663.80.3.366
- Marsh, H. W., Byrne, B. M. & Shavelson, R. J. (1992). A multidimensional, hierarchical self-concept: Theoretical and empirical justification. In *The self: Definitional and methodological issues* (S. 44–95). SUNY series, studying the self. State University of New York Press. doi:10.1007/BF01322177
- Marsh, H. W., Byrne, B. M. & Yeung, A. S. (1999). Causal ordering of academic self-concept and achievement: Reanalysis of a pioneering study and revised recommendations. *Educational Psychologist*, *34*(3), 155–167. doi:10.1207/s15326985ep3403_2
- Marsh, H. W., Craven, R. & Debus, R. (1998). Structure, stability, and development of young children's self-concepts: A multicohort–multioccasion study. *Child Development*, *69*(4), 1030–1053. doi:10.1111/j.1467-8624.1998.tb06159.x
- Marsh, H. W., Parker, J. & Barnes, J. (1985). Multidimensional adolescent self-concepts: Their relationship to age, sex, and academic measures. *American Educational Research Journal*, *22*(3), 422–444. doi:10.3102/00028312022003422
- Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O. & Baumert, J. (2005b). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects mo-

- dels of causal ordering. *Child Development*, 76(2), 397–416. doi:10.1111/j.1467-8624.2005.00853.x
- Marsh, H. W., Walker, R. & Debus, R. (1991). Subject-specific components of academic self-concept and self-efficacy. *Contemporary Educational Psychology*, 16(4), 331–345. doi:10.1016/0361-476X(91)90013-B
- Marsh, H. W. & Yeung, A. S. (1997). Causal effects of academic self-concept on academic achievement: Structural equation models of longitudinal data. *Journal of Educational Psychology*, 89(1), 41–54. doi:10.1037/0022-0663.89.1.41
- Marthaler, D. (2007). *Testen kognitiver Fähigkeiten per Telefon*, Universität des Saarlandes. Saarbrücken: Unveröffentlichte Diplomarbeit.
- Martin, A. J., Anderson, J., Bobis, J., Way, J. & Vellar, R. (2012). Switching on and switching off in mathematics: An ecological study of future intent and disengagement among middle school students. *Journal of Educational Psychology*, 104(1), 1–18. doi:10.1037/a0025988
- McClearn, G. E., Johansson, B., Berg, S., Pedersen, N. L., Ahern, F., Petrill, S. A. & Plomin, R. (1997). Substantial genetic influence on cognitive abilities in twins 80 or more years old. *Science*, 276(5318), 1560–1563. doi:10.1126/science.276.5318.1560
- McClelland, D. C., Atkinson, J. W., Clark, R. A. & Lowell, E. L. (1953). *The achievement motive*. New York: Appleton-Century-Crofts.
- McDougall, W. (1937). *Aufbaukräfte der Seele*. Leipzig: Thieme.
- McFarland, C. & Buehler, R. (1995). Collective self-esteem as a moderator of the frog-pond effect in reactions to performance feedback. *Journal of Personality and Social Psychology*, 68(6), 1055–1070. doi:10.1037/0022-3514.68.6.1055
- McGue, M., Bouchard, T. J., Iacono, W. G. & Lykken, D. T. (1993). Behavioral genetics of cognitive ability: A life-span perspective. In R. Plomin & G. E. McClearn (Hrsg.), *Nature, nurture and psychology* (S. 59–76). Washington, DC: American Psychological Association. doi:10.1037/10131-003
- McGue, M. & Christensen, K. (2001). The heritability of cognitive functioning in very old adults: Evidence from Danish twins aged 75 years and older. *Psychology and Aging*, 16(2), 272–280. doi:10.1037/0882-7974.16.2.272
- Mitchell, K. S., Mazzeo, S. E., Bulik, C. M., Aggen, S. H., Kendler, K. S. & Neale, M. C. (2007). An investigation of a measure of twins' equal environments. *Twin Research and Human Genetics*, 10(6), 840–847. doi:10.1375/twin.10.6.840
- Moffitt, T. E., Caspi, A. & Rutter, M. (2005). Strategy for investigating interactions between measured genes and measured environments. *Archives of General Psychiatry*, 62(5), 473–481.
- 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 & Psychiatry & Allied Disciplines*, 34(4), 455–506. doi:10.1111/j.1469-7610.1993.tb01031.x
- Möller, J. & Trautwein, U. (2009). Selbstkonzept. In *Pädagogische Psychologie* (Kap. 8, S. 179–203). Heidelberg: Springer. doi:10.1007/978-3-540-88573-3_8
- Möller, J. & Köller, O. (2004). Die Genese akademischer Selbstkonzepte: Effekte dimensionaler und sozialer Vergleiche. *Psychologische Rundschau*, 55(1), 19–27. doi:10.1026/0033-3042.55.1.19
- Morris-Yates, A., Andrews, G., Howie, P. & Henderson, S. (1990). Twins: A test of the equal environments assumption. *Acta Psychiatrica Scandinavica*, 81(4), 322–326. doi:10.1111/j.1600-0447.1990.tb05457.x
- Moschner, B. & Dickhäuser, O. (2006). Selbstkonzept. In D. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 685–692). Weinheim.
- Muris, P., Meesters, C. & Fijen, P. (2003). The self-perception profile for children: Further evidence for its factor structure, reliability, and validity. *Personality and Individual Differences*, 35(8), 1791–1802. doi:10.1016/S0191-8869(03)00004-7
- Murphy, P. K. & Alexander, P. A. (2000). A motivated exploration of motivation terminology. *Contemporary Educational Psychology*, 25(1), 3–53. doi:10.1006/ceps.1999.1019
- Murray, H. A. (1938). *Explorations in personality*. New York: Oxford University Press.
- Muthén, L. K. & Muthén, B. O. (1998-2010). *Mplus Users Guide. Sixth Edition*. Los Angeles, CA: Muthén & Muthén.
- Muthén, L. K. & Weber, C. (29. Januar 2010). Chi-Square Diff Testing Using the Satorra-Bentler Scaled Chi-Square. Zugriff am 29. Juni 2012, unter <http://www.statmodel.com/discussion/messages/9/189.html1332955599>
- Nagengast, B. & Marsh, H. W. (2011). The negative effect of school-average ability on science self-concept in the UK, the UK countries and the world: The Big-Fish-Little-Pond-Effect for PISA 2006. *Educational Psychology*, 31(5), 629–656. doi:10.1080/01443410.2011.586416
- Neale, M. C. (2009). Biometrical models in behavioral genetics. In Y. K. Kim (Hrsg.), *Handbook of Behavioral Genetics* (S. 15–34). Springer Verlag. doi:10.1007/978-0-387-76727-7_2
- Neale, M. C., Boker, S. M., Bergeman, C. S. & Maes, H. H. (2006). The utility of genetically informative data in the study of development. In C. S. Bergeman & S. M. Boker (Hrsg.), *Methodological issues in aging research* (S. 269–327). London: Lawrence Erlbaum Associates.
- Neale, M. C., Boker, S. M., Xie, G. & Maes, H. H. (2003). Mx: Statistical modeling. Zugriff am 29. April 2012, unter <http://www.vcu.edu/mx/documentation.html>
- Neale, M. C. & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Dordrecht: Kluwer Academic Publishers. doi:10.1007/BF01067561

- Neale, M. C., Lubke, G., Aggen, S. H. & Dolan, C. V. (2005). Problems with using sum scores for estimating variance components: Contamination and measurement noninvariance. *Twin Research and Human Genetics*, *8*(6), 553–568. doi:10.1375/183242705774860231
- Neiderhiser, J. M., Reiss, D., Pedersen, N. L., Lichtenstein, P., Spotts, E. L., Hansson, K., ... Elthammer, O. (2004). Genetic and environmental influences on mothering of adolescents: A comparison of two samples. *Developmental psychology*, *40*(3), 335–351. doi:10.1037/0012-1649.40.3.335
- Neiss, M. B., Sedikides, C. & Stevenson, J. (2002). Self-esteem: A behavioural genetic perspective. *European Journal of Personality*, *16*(5), 351–367. doi:10.1002/per.456
- Neisser, U., Boodoo, G., Bouchard, J. T. J., Boykin, A. W., Brody, N., Ceci, S. J., ... Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*(2), 77–101. doi:10.1037/0003-066X.51.2.77
- Neubauer, A. C., Spinath, F. M., Riemann, R., Borkenau, P. & Angleitner, A. (2000). Genetic and environmental influences on two measures of speed of information processing and their relation to psychometric intelligence: Evidence from the German observational study of adult twins. *Intelligence*, *28*(4), 267–289.
- Newman, R. S. (1984). Children's achievement and self-evaluations in mathematics: A longitudinal study. *Journal of Educational Psychology*, *76*(5), 857–873. doi:10.1037/0022-0663.76.5.857
- Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F. & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*. doi:10.1037/a0026699
- Nurmi, A.-E. & Aunola, K. (2005). Task-motivation during the first school years: A person-oriented approach to longitudinal data. *Learning and Instruction*, *15*(2), 103–122. doi:10.1016/j.learninstruc.2005.04.009
- Obach, M. S. (2003). A longitudinal-sequential study of perceived academic competence and motivational beliefs for learning among children in middle school. *Educational Psychology*, *23*(3), 323–338. doi:10.1080/0144341032000060156
- Oliver, B. R. & Plomin, R. (2007). Twins' Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behavior problems from childhood through adolescence. *Twin Research and Human Genetics*, *10*(1), 96–105. doi:10.1375/twin.10.1.96
- Ozaki, K., Toyoda, H., Iwama, N., Kubo, S. & Ando, J. (2011). Using non-normal SEM to resolve the ACDE model in the classical twin design. *Behavior genetics*, *41*(2), 329–339. doi:DOI:10.1007/s10519-010-9386-5
- Pallant, J. (2010). *SPSS survival manual: A step by step guide to data analysis using SPSS* (4th). Berkshire: McGraw Hill Education.

- Pedersen, N. L. & McClearn, G. E. (1994). Is there G beyond g (Is there genetic influence on specific cognitive abilities independent of genetic influence on general cognitive ability). *Intelligence*, 18(2), 133–143.
- Pedersen, N. L., Plomin, R., Nesselroade, J. R. & McClearn, G. E. (1992). A quantitative genetic analysis of cognitive abilities during the second half of the life span. *Psychological Science*, 3(6), 346–353. doi:10.1111/j.1467-9280.1992.tb00045.x
- Pedhazur, E. J. (1982). *Multiple regression in behavioral research*. New York: Holt, Rinehart und Winston.
- Petrill, S. A., Deater-Deckard, K., Thompson, L. A., De Thorne, L. S. & Schatschneider, C. (2006). Reading skills in early readers. Genetic and shared environmental influences. *Journal of Learning Disabilities*, 39(1), 48. doi:10.1177/00222194060390010501
- Petrill, S. A., Lipton, P. A., Hewitt, J. K., Plomin, R., Cherny, S. S., Corley, R. & DeFries, J. C. (2004). Genetic and environmental contributions to general cognitive ability through the first 16 years of life. *Developmental Psychology*, 40(5), 805–812. doi:10.1037/0012-1649.40.5.805
- Petrill, S. A., Rempell, J., Oliver, B. & Plomin, R. (2002). Testing cognitive abilities by phone in a sample of 6-to 8-year-olds. *Intelligence*, 30(4), 353–360. doi:10.1016/S0160-2896(02)00087-9
- Petrill, S. A., Saudino, K. S., Wilkerson, B. & Plomin, R. (2001). Genetic and environmental molarity and modularity of cognitive functioning in 2-year-old twins. *Intelligence*, 29(1), 31–43. doi:10.1016/S0160-2896(00)00041-6
- Plomin, R. & Daniels, D. (1987). Why are children in the same family so different from one another. *Behavioral and Brain Sciences*, 10(01), 1–16. doi:http://dx.doi.org/10.1017/S0140525X00055941
- Plomin, R. & DeFries, J. C. (1998). The genetics of cognitive abilities and disabilities. *Scientific American*, 278(5), 62–69.
- Plomin, R., DeFries, J. C. & Loehlin, J. C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin*, 84(2), 303–322. doi:10.1037/0033-2909.84.2.309
- Plomin, R., DeFries, J. C., McClearn, G. E. & McGuffin, P. (2008). *Behavioral genetics*. New York: Worth Publishers.
- Plomin, R., Fulker, D. W., Corley, R. & DeFries, J. C. (1997). Nature, nurture, and cognitive development from 1 to 16 years: A parent-offspring adoption study. *Psychological Science*, 8(6), 442–447. doi:10.1111/j.1467-9280.1997.tb00458.x
- Plomin, R. & Spinath, F. M. (2002). Genetics and general cognitive ability (g). *Trends in Cognitive Sciences*, 6(4), 169–176. doi:10.1016/S1364-6613(00)01853-2

- Plomin, R. & Spinath, F. M. (2004). Intelligence: Genetics, genes, and genomics. *Journal of Personality and Social Psychology*, *86*(1), 112–129. doi:10.1037/0022-3514.86.1.112
- Plomin, R., Asbury, K. & Dunn, J. (2001). Why are children in the same family so different? Nonshared environment a decade later. *Canadian Journal of Psychiatry*, *46*(3), 225–233.
- Plomin, R. & Caspi, A. (1998). DNA and personality. *European Journal of Personality*, *12*(5), 387–407. doi:10.1002/(SICI)1099-0984(199809)12:5<387::AID-PER330>3.0.CO;2-U
- Pooch, Y. (2006). *Interne und externe Validität: Fernerhebung von Intelligenz neunjähriger Grundschüler*, Universität des Saarlandes. Saarbrücken: Unveröffentlichte Diplomarbeit.
- Posthuma, D. (2009). Multivariate genetic analysis. In Y. K. Kim (Hrsg.), *Handbook of Behavior Genetics* (S. 47–59). New York: Springer Verlag. doi:10.1007/978-0-387-76727-7_4
- Posthuma, D., De Geus, E. J. C., Bleichrodt, N. & Boomsma, D. I. (2000). Twin-singleton differences in intelligence. *Twin Research and Human Genetics*, *3*(2), 83–87. doi:10.1375/136905200320565535
- Posthuma, D., De Geus, E. J. C. & Boomsma, D. I. (2001). Perceptual speed and IQ are associated through common genetic factors. *Behavior Genetics*, *31*(6), 593–602. doi:10.1023/A:1013349512683
- Posthuma, D. & De Geus, E. (2006). Progress in the molecular-genetic study of intelligence. *Current Directions in Psychological Science*, *15*(4), 151–155. doi:10.1111/j.1467-8721.2006.00426.x
- Preckel, F. & Brüll, M. (2010). The benefit of being a big fish in a big pond: Contrast and assimilation effects on academic self-concept. *Learning and Individual Differences*, *20*(5), 522–531. doi:10.1016/j.lindif.2009.12.007
- Price, T. S., Eley, T. C., Dale, P. S., Stevenson, J., Saudino, K. & Plomin, R. (2000a). Genetic and environmental covariation between verbal and nonverbal cognitive development in infancy. *Child Development*, *71*(4), 948–959. doi:10.1111/1467-8624.00201
- Price, T. S., Freeman, B., Craig, I., Petrill, S. A., Ebersole, L. & Plomin, R. (2000b). Infant zygosity can be assigned by parental report questionnaire data. *Twin Research And Human Genetics*, *3*(3), 129–133. doi:10.1375/136905200320565391
- Purcell, S. (2002). Variance components models for gene-environment interaction in twin analysis. *Twin Research*, *5*(6), 554–571. doi:10.1375/136905202762342026
- Ratelle, C. F., Guay, F., Vallerand, R. J., Larose, S. & Senécal, C. (2007). Autonomous, controlled, and amotivated types of academic motivation: A person-oriented ana-

- lysis. *Journal of Educational Psychology*, 99(4), 734–746. doi:10.1037/0022-0663.99.4.734
- Reinders, H. (2006). Kausalanalysen in der Längsschnittforschung. Das Crossed-Lagged-Panel Design. *Diskurs Kindheits- und Jugendforschung*, 4, 569–587.
- Reinecke, J. (2005). *Strukturgleichungsmodelle in den Sozialwissenschaften*. München: Oldenbourg Verlag.
- Reusser, K. (1997). Erwerb mathematischer Kompetenzen: Literaturüberblick. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 141–155). Weinheim: Beltz/PVU.
- Reyes, M. R., Brackett, M. A., Rivers, S. E., White, M. & Salovey, P. (2012). Classroom emotional climate, student engagement, and academic achievement. *Journal of Educational Psychology*, 104(3), 700–712. doi:10.1037/a0027268
- Reynolds, C. A., Finkel, D., McArdle, J. J., Berg, S., Pedersen, N. L. & Gatz, M. (2005). Quantitative genetic analysis of latent growth curve models of cognitive abilities in adulthood. *Developmental psychology*, 41(1), 3–16. doi:10.1037/0012-1649.41.1.3
- Reznick, J. S., Corley, R., Robinson, J. & Matheny, A. P. (1997). A longitudinal twin study of intelligence in the second year. *Monographs of the Society for Research in Child Development*, 62(249), 1–154.
- Rheinberg, F. (2006). *Motivation*. Stuttgart: Kohlhammer.
- Rheinberg, F. & Krug, S. (1999). *Motivationsförderung im Schulalltag*. Hogrefe.
- Riemann, R. & Spinath, F. M. (2005). Genetik der Persönlichkeit. In J. Hennig & P. Netter (Hrsg.), *Biopsychologische Grundlagen der Persönlichkeit* (S. 539–628). Heidelberg: Spektrum Akademischer Verlag.
- Rietveld, M. J. H., Dolan, C. V., vanBaal, G. C. M. & Boomsma, D. I. (2003). A twin study of differentiation of cognitive abilities in childhood. *Behavior Genetics*, 33(4), 367–381. doi:10.1023/A:1025388908177
- Rietz, C. & Rudinger, G. (2007). Analyse von Längsschnittdaten. In M. Hasselhorn & W. Schneider (Hrsg.), *Handbuch der Entwicklungspsychologie* (S. 635–645). Göttingen: Hogrefe.
- Rijsdijk, F. V. & Sham, P. C. (2002). Analytic approaches to twin data using structural equation models. *Briefings in bioinformatics*, 3(2), 119–133. doi:10.1093/bib/3.2.119
- Rijsdijk, F. V., Vernon, P. A. & Boomsma, D. I. (2002). Application of hierarchical genetic models to Raven and WAIS subtests: A Dutch twin study. *Behavior Genetics*, 32(3), 199–210. doi:10.1023/A:1016021128949
- Rindermann, H. (2006). Was messen internationale Schulleistungstudien? Schulleistungen, Schülerfähigkeiten, kognitive Fähigkeiten, Wissen oder allgemeine Intelligenz. *Psychologische Rundschau*, 57(2), 69–86. doi:10.1026/0033-3042.57.2.69

- Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R. & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes. *Psychological Bulletin*, *130*(2), 261–288. doi:10.1037/0033-2909.130.2.261
- Roth, B., Becker, N., Spengler, M. & Spinath, F. M. (2011). *Die Validität von Intelligenz bei der Vorhersage von Schulnoten - eine Metaanalyse. Vortrag auf der 11. Arbeitstagung der Fachgruppe Differentielle Psychologie, Persönlichkeitspsychologie und Psychologische Diagnostik der Deutschen Gesellschaft für Psychologie, Saarbrücken, 26.09.2011 - 28.09.2011.*
- Rothermund, K. (2003). Analyse längsschnittlicher Veränderungen in Strukturgleichungsmodellen. *Trierer Psychologische Berichte*, *30*(1), 1–22.
- Roy, M. A., Neale, M. C. & Kendler, K. S. (1995). The genetic epidemiology of self-esteem. *The British Journal of Psychiatry*, *166*(6), 813–820. doi:10.1192/bjp.166.6.813
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. doi:10.1093/biomet/63.3.581
- Rutter, M. (2006). *Genes and behavior: Nature-nurture interplay explained.* Oxford: Blackwell Publishing. doi:10.1111/j.1601-183X.2006.00219.x
- Rutter, M., Pickles, A., Murray, R. & Eaves, L. (2001). Testing hypotheses on specific environmental causal effects on behavior. *Psychological Bulletin*, *127*(3), 291–324. doi:10.1037/0033-2909.127.3.291
- Ryan, R. M. & Deci, E. L. (2000a). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54–67. doi:10.1006/ceps.1999.1020
- Ryan, R. M. & Deci, E. L. (2000b). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. doi:10.1037/0003-066X.55.1.68
- Satorra, A. & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*(4), 507–514. doi:10.1007/BF02296192
- Saudino, K. J., McGuire, S., Reiss, D., Hetherington, E. M. & Plomin, R. (1995). Parent ratings of EAS temperaments in twins, full siblings, half siblings, and step siblings. *Journal of Personality and Social Psychology*, *68*(4), 723–733. doi:10.1037/0022-3514.68.4.723
- Sauer, J. & Gamsjäger, E. (1996). *Ist Schulerfolg vorhersagbar? Die Determinanten der Grundsulleistung und ihr prognostischer Wert für den Sekundarschulerfolg.* Göttingen: Hogrefe.
- Sauer, J. & Gattringer, H. (1986). Zur Aufklärung der Schulleistung durch spezifische und gemeinsame Varianzanteile von Intelligenz und Motivation. In K. Daumenlang & J. Sauer (Hrsg.), *Aspekte psychologischer Forschung. Festschrift zum 60. Geburtstag von Erwin Roth* (S. 237–257). Göttingen: Hogrefe.

- Scarr, S. & McCartney, K. (1983). How people make their own environments: A theory of genotype–environment effects. *Child development*, 424–435.
- Schaie, K. W. (1996). Intellectual development in adulthood. In J. E. Birren & K. W. Schaie (Hrsg.), *Handbook of the psychology of aging* (S. 266–286). San Diego: CA: Academic Press.
- Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23–74.
- Schicke, M. C. & 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(1), 62–69.
- Schiefele, U. (2009). Motivation. In J. Möller & E. Wild (Hrsg.), *Pädagogische Psychologie* (Kap. 7, S. 151–177). Heidelberg: Springer.
- Schiefele, U. & Köller, O. (2006). Intrinsische und extrinsische Motivation. In D. Rost (Hrsg.), *Handbuch Pädagogische Psychologie* (S. 303–310). Weinheim: Beltz.
- Schiefele, U. & Schreyer, I. (1994). Intrinsische Lernmotivation und Lernen. Ein Überblick zu Ergebnissen aus der Forschung. *Zeitschrift für Pädagogische Psychologie*, 8(1), 1–13.
- Schiefele, U. (1999). Interest and learning from text. *Scientific Studies of Reading*, 3(3), 257–279. doi:10.1207/s15327999xssr0303_4
- Schiefele, U., Krapp, A. & Schreyer, I. (1993). Metaanalyse des Zusammenhangs von Interesse und schulischer Leistung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 25(2), 120–148.
- Schmid, M. (1998). Entwicklungsaufgaben. In H. Häcker & K. Stapf (Hrsg.), *Dorsch Psychologisches Wörterbuch* (13. überarb. und erw. Aufl, S. 228). Bern: Verlag Hans Huber.
- Schmiedek, F. & Lindenberger, U. (2007). Methodologische Grundlagen. In J. Brandtstädter & U. Lindenberger (Hrsg.), *Entwicklungspsychologie der Lebensspanne* (1. Aufl., Kap. 1.3, S. 67–96). Stuttgart: Kohlhammer.
- Schmitt, M. (2003). Stabilität. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik*. Weinheim: Beltz.
- Schmitz, S., Cherny, S. S. & Fulker, D. W. (1998). Increase in power through multivariate analyses. *Behavior Genetics*, 28(5), 357–363. doi:10.1023/A:1021669602220
- Schneider, W. (1991). Methodische Probleme und Möglichkeiten schulbezogener Längsschnittforschung. In H. Pekrun R. und Fend (Hrsg.), *Schule und Persönlichkeitsentwicklung: Ein Resümee der Längsschnittforschung* (S. 57–80). Stuttgart: Enke.
- Schneider, W. & Bullock, M. (2008). *Human development from early childhood to early adulthood. Findings from a 20 year longitudinal study*. New York: Psychology Press.

- Schneider, W. & Stefanek, J. (2004). Entwicklungsveränderungen allgemeiner kognitiver Fähigkeiten und schulbezogener Fertigkeiten im Kindes- und Jugendalter. Evidenz für einen Schereneffekt. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 36(3), 147–159. doi:10.1026/0049-8637.36.3.147
- Schneider, W., Stefanek, J. & Dotzler, H. (1997). Erwerb des Lesens und des Recht Schreibens: Ergebnisse aus dem SCHOLASTIK-Projekt. In F. E. Weinert & A. Helmke (Hrsg.), *Leistungsmessungen in Schulen* (S. 113–134). Weinheim: Beltz.
- Schöne, C., Dickhäuser, O., Spinath, B. & Stiensmeier-Pelster. (2003). Das Fähigkeitsselbstkonzept und seine Erfassung. In J. Stiensmeier-Pelster & F. Rheinberg (Hrsg.), *Diagnostik von Motivation und Selbstkonzept. Jahrbuch der pädagogisch-psychologischen Diagnostik. Tests und Trends* (S. 3–14). Neue Folge, Band 2. Bern: Hogrefe.
- Schulenberg, J. E., Asp, C. E. & Petersen, A. C. (1984). School from the young adolescent's perspective: A descriptive report. *Journal of Early Adolescence*, 4(2), 107–130. doi:10.1177/0272431684042002
- Schunk, D. H., Pintrich, P. R. & Meece, J. L. (2008). *Motivation in education*. Upper Saddle River, NJ: Pearson Education.
- 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. doi:10.1037/h0084155
- Schwarzer, R., Lange, B. & Jerusalem, M. (1982). Selbstkonzeptentwicklung nach einem Bezugsgruppenwechsel. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 14(2), 125–140.
- Scottish Council for Research in Education. (1933). *The intelligence of Scottish children: A national survey of an age-group*. London: University of London Press.
- Segal, N. L. & Johnson, W. (2009). Twin studies of general mental ability. In Y. K. Kim (Hrsg.), *Handbook of Behavior Genetics* (S. 81–99). New York: Springer. doi:10.1007/978-0-387-76727-7_6
- Seginer, R. (2006). Parents' educational involvement: A developmental ecology perspective. *Parenting: Science and Practice*, 6(1), 1–48. doi:10.1207/s15327922par0601_1
- Seidman, E., Allen, L., Aber, J. L., Mitchell, C. & Feinman, J. (1994). The impact of school transitions in early adolescence on the self-system and perceived social context of poor urban youth. *Child Development*, 56(2), 507–522. doi:10.1111/j.1467-8624.1994.tb00766.x
- Shanahan, M. J., Hofer, S. M. & Shanahan, L. (2003). Biological models of behavior and the life course. In J. T. Mortimer & M. J. Shanahan (Hrsg.), *Handbook of the life course* (S. 597–622). New York: Kluwer Academic.

- Shanahan, M. J., Vaisey, S., Erickson, L. D. & Smolen, A. (2008). Environmental contingencies and genetic propensities: Social capital, educational continuation, and dopamine receptor gene DRD2. *American Journal of Sociology*, 114(1), 260–286. doi:10.1086/592204
- Shavelson, R. J., Hubner, J. J. & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, 46(3), 407–441. doi:10.3102/00346543046003407
- Sirsch, U. (2000). *Probleme beim Schulwechsel. Die subjektive Bedeutung des bevorstehenden Wechsels von der Grundschule in die weiterführende Schule*. Dissertation. Münster: Waxmann.
- Skaalvik, E. M. & Rankin, R. J. (1995). A test of the internal/external frame of reference model at different levels of math and verbal self-perception. *American Educational Research Journal*, 32(1), 161–184. doi:10.3102/00028312032001161
- 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(2), 135–149. doi:10.1080/00220979909598349
- Skaalvik, E. M. & Hagtvet, K. A. (1990). Academic achievement and self-concept: An analysis of causal predominance in a developmental perspective. *Journal of Personality and Social Psychology*, 58(2), 292–307. doi:10.1037/0022-3514.58.2.292
- 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(3), 1187–1202. doi:10.2466/pr0.95.3f.1187-1202
- Sontag, L. W., Baker, C. T. & Nelson, V. L. (1958). Mental growth and personality development: A longitudinal study. *Monographs of the Society for Research in Child Development*, 23(2), 1–143.
- Spearman, C. E. (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* (Diss., Universität des Saarlandes). Zugriff am 22. April 2012, unter <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., Freudenthaler, H. & Neubauer, A. C. (2010). Domain-specific school achievement in boys and girls as predicted by intelligence, personality and motivation.

- Personality and Individual Differences*, 48(4), 481–486. doi:10.1016/j.paid.2009.11.028
- Spinath, B. & Spinath, F. M. (2005a). Development of self-perceived ability in elementary school: The role of parents' perceptions, teacher evaluations, and intelligence. *Cognitive Development*, 20(2), 190–204. doi:10.1016/j.cogdev.2005.01.001
- Spinath, B. & Spinath, F. M. (2005b). Longitudinal analysis of the link between learning motivation and competence beliefs among elementary school children. *Learning and Instruction*, 15(2), 87–102. doi:10.1016/j.learninstruc.2005.04.008
- 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(4), 363–374. doi:10.1016/j.intell.2005.11.004
- 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. doi:10.1111/j.1467-8624.2008.01205.x
- Spinath, F. M. & Deary, I. J. (2008). Verhaltensgenetik der Intelligenz. In F. J. Neyer & F. M. Spinath (Hrsg.), *Anlage und Umwelt* (S. 129–147). Stuttgart: Lucius & Lucius.
- Spinath, F. M. & Johnson, W. (2011). Behavior genetics. In T. Chamorro-Premuzic, S. von Stumm & A. Furnham (Hrsg.), *The Wiley-Blackwell Handbook of Individual Differences* (S. 271–304). Chichester: John Wiley & Sons, Ltd. doi:10.1002/9781405184359.ch10
- Spinath, F. M., Ronald, A., Harlaar, N., Price, T. S. & Plomin, R. (2003). Phenotypic g early in life: On the etiology of general cognitive ability in a large population sample of twin children aged 2–4 years. *Intelligence*, 31(2), 195–210. doi:10.1016/S0160-2896(02)00110-1
- Spinath, F. M., Spinath, B. & Plomin, R. (2008a). The nature and nurture of intelligence and motivation in the origins of sex differences in elementary school achievement. *European Journal of Personality*, 22(3), 211–229. doi:10.1002/per.677
- Spinath, F. M., Toussaint, A., Spengler, M. & Spinath, B. (2008b). Motivation als Element schulbezogener Selbstregulation: Die Rolle genetischer Einflüsse. *Unterrichtswissenschaft*, 36(1), 3–16.
- Spinath, F. M. & Wolf, H. (2006). CoSMoS and TwinPaW: Initial report on two new German twin studies. *Twin Research and Human Genetics*, 9(6), 787–790. doi:10.1375/183242706779462903
- Statistisches Bundesamt. (2010). *Statistisches Jahrbuch 2010. Für die Bundesrepublik Deutschland mit 'Internationalen Übersichten'*. Wiesbaden: Statistisches Bundesamt.

- Steinmayr, R., Bipp, T. & Spinath, B. (2011). Goal orientations predict academic performance beyond intelligence and personality. *Learning and Individual Differences*, *21*(2), 196–200. doi:10.1016/j.lindif.2010.11.026
- Steinmayr, R. & Spinath, B. (2008). Sex differences in school achievement: What are the roles of personality and achievement motivation. *European Journal of Personality*, *22*(3), 185–209. doi:10.1002/per.676
- Steinmayr, R. & Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning and Individual Differences*, *19*(1), 80–90. doi:10.1016/j.lindif.2008.05.004
- Stern, E. (26. Juni 2003). Wissen schlägt Intelligenz/Interviewer: T. Kerstan. Zugriff am 22. April 2012, unter http://scienceblogs.com/pharyngula/2007/01/the_unfortunate_prerequisites.php
- Sternberg, R. J. (2000). *Handbook of Intelligence*. Cambridge: Cambridge University Press.
- Sternberg, R. J. (2003). Our research program validating the triarchic theory of successful intelligence: Reply to Gottfredson. *Intelligence*, *31*(4), 399–413. doi:10.1016/S0160-2896(02)00143-5
- Sternberg, R. J. (2004). Culture and intelligence. *American Psychologist*, *59*(5), 325–338. doi:10.1037/0003-066X.59.5.325
- Sternberg, R. J. (2005). The triarchic theory of successful intelligence. In P. D. Flanagan & P. L. Harrison (Hrsg.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (S. 103–119). New York: Guilford Press.
- Sternberg, R. J. & Detterman, D. K. (1986). *What is intelligence: Contemporary viewpoints on its nature and definition*. Norwood: Ablex Publishing Corporation.
- Stipek, D. J. & Daniels, D. H. (1988). Declining perceptions of competence: A consequence of changes in the child or in the educational environment. *Journal of Educational Psychology*, *80*(3), 352–356. doi:10.1037/0022-0663.80.3.352
- Stipek, D. & Iver, D. (1989). Developmental change in children's assessment of intellectual competence. *Child Development*, *60*(3), 521–538. doi:10.2307/1130719
- Stromswold, K. (2001). The heritability of language: A review and metaanalysis of twin, adoption, and linkage studies. *Language*, *77*(4), 647–723.
- Tabachnick, B. & Fidell, L. (2001). *Using multivariate statistics* (4th). Boston: Allyn & Bacon.
- Tent, L. (2006). Zensuren. In D. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 873–880). Weinheim: Beltz.
- Tewes, U., Rossmann, P. & Schallberger, U. (2000). *Hamburg-Wechsler-Intelligenztest für Kinder III*. Bern: Huber.
- Thompson, L. A., Detterman, D. K. & Plomin, R. (1991). Associations between cognitive abilities and scholastic achievement: Genetic overlap but environmental

- differences. *Psychological Science*, *2*(3), 158–165. doi:10.1111/j.1467-9280.1991.tb00124.x
- Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.
- Toussaint, A. (2007). *Inkrementeller Beitrag der Leistungsmotivation bei der Vorhersage von schulischem Erfolg im Grundschulalter: Eine Zwillingsstudie*. Unveröffentlichte Diplomarbeit, Universität des Saarlandes. Saarbrücken.
- Trautwein, U., Lüdtke, O., Köller, O. & Baumert, J. (2006a). Self-esteem, academic self-concept, and achievement: How the learning environment moderates the dynamics of self-concept. *Journal of Personality and Social Psychology*, *90*(2), 334–349. doi:10.1037/0022-3514.90.2.334
- Trautwein, U., Lüdtke, O., Marsh, H., Köller, O. & Baumert, J. (2006b). Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology*, *98*(4), 788–806. doi:10.1037/0022-0663.98.4.788
- Trautwein, U. & Lüdtke, O. (2005). The big-fish-little-pond effect. Future research questions and educational implications. *Zeitschrift für Pädagogische Psychologie*, *19*(3), 137–140. doi:10.1024/1010-0652.19.3.137
- Trouton, A., Spinath, F. M. & Plomin, R. (2002). Twins Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behavior problems in childhood. *Twin Research*, *5*(5), 444–448. doi:10.1375/136905202320906255
- Tucker-Drob, E. M., Rhemtulla, M., Harden, K. P., Turkheimer, E. & Fask, D. (2011). Emergence of a gene x socioeconomic status interaction on infant mental ability between 10 months and 2 years. *Psychological Science*, *22*(1), 125–133. doi:10.1177/0956797610392926
- Turkheimer, E., Haley, A., Waldron, M., D’Onofrio, B. & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science*, *14*(6), 623–628. doi:10.1046/j.0956-7976.2003.psci_1475.x
- Turkheimer, E., D’Onofrio, B. M., Maes, H. H. & Eaves, L. J. (2005). Analysis and interpretation of twin studies including measures of the shared environment. *Child Development*, *76*(6), 1217–1233. doi:10.1111/j.1467-8624.2005.00845.x-i1
- Turkheimer, E. & Waldron, M. (2000). Nonshared environment: A theoretical, methodological, and quantitative review. *Psychological Bulletin*, *126*(1), 78–108. doi:10.1037/0033-2909.126.1.78
- Valentine, J. C., DuBois, D. L. & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*, *39*(2), 111–133. doi:10.1207/s15326985ep3902_3
- van der Sluis, S., Willemsen, G., de Geus, E. J. C., Boomsma, D. I. & Posthuma, D. (2008). Gene-environment interaction in adults’ IQ scores: Measures of past and

- present environment. *Behavior Genetics*, 38(4), 348–360. doi:10.1007/s10519-008-9212-5
- Vernon, P. E. (1964). *The structure of human abilities* (2. Aufl.). New York: Wiley.
- Vinkhuyzen, A. A. E., van der Sluis, S., Maes, H. H. & Posthuma, D. (2012). Reconsidering the heritability of intelligence in adulthood: Taking assortative mating and cultural transmission into account. *Behavior Genetics*, 1–12. doi:10.1007/s10519-011-9507-9
- Visscher, P. M. (2004). Power of the classical twin design revisited. *Twin Research*, 7(5), 505–512.
- Visscher, P. M., Gordon, S. & Neale, M. C. (2008). Power of the classical twin design revisited: II Detection of common environmental variance. *Twin Research and Human Genetics*, 11(1), 48–54.
- 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.
- Wainwright, M. A., Wright, M. J., Luciano, M., Geffen, G. M. & Martin, N. G. (2005). Multivariate genetic analysis of academic skills of the Queensland core skills test and IQ highlight the importance of genetic g. *Twin Research and Human Genetics*, 8(6), 602–608. doi:10.1375/183242705774860259
- Walker, S. O., Petrill, S. A. & Plomin, R. (2005). A genetically sensitive investigation of the effects of the school environment and socio-economic status on academic achievement in seven-year-olds. *Educational Psychology*, 25(1), 55–73.
- Walker, S. O., Petrill, S. A., Spinath, F. M. & Plomin, R. (2004). Nature, nurture and academic achievement: A twin study of teacher assessments of 7-year-olds. *British Journal of Educational Psychology*, 74(3), 323–342. doi:10.1348/0007099041552387
- Walker, S. O. & Plomin, R. (2006). Nature, nurture, and perceptions of the classroom environment as they relate to teacher-assessed academic achievement: A twin study of nine-year-olds. *Educational Psychology*, 26(4), 541–561. doi:10.1080/01443410500342500
- Wang, M. C., Haertel, G. D. & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, 63(3), 249–294. doi:10.3102/00346543063003249
- Weber, H. S. (2012). *Die relative Bedeutung kognitiver und non-kognitiver Bedingungsfaktoren bei der Vorhersage des Schulerfolgs: Vergleich der prädiktiven Validität von Arbeitsgedächtnis, Intelligenz, Motivation und SES* (Diss., Universität des Saarlandes). Zugriff am 22. April 2012, unter <http://scidok.sulb.uni-saarland.de/volltexte/2012/4900/>

- Wechsler, D. (1944). *The measurement of adult intelligence*. Baltimore: Williams & Wilkins. doi:10.1037/11329-009
- Wechsler, D. (1992). *Wechsler Intelligence Scale for Children Third Edition UK (WISC-IIIUK) Manual*. London: The Psychological Corporation.
- Weiber, R. S. & Mühlhaus, D. D. (2010). *Strukturgleichungsmodellierung – Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS*. Heidelberg: Springer.
- Weinert, F. E. (2002). Vorwort. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (2. Auflage, S. 5–7). Weinheim: Beltz.
- Weinert, F. E. & Stefanek, J. (1997). Entwicklung vor, während und nach der Grundschulzeit: Ergebnisse aus dem SCHOLASTIK-Projekt. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 423–451). Weinheim: Beltz.
- West, S. G., Finch, J. F. & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Hrsg.), *Structural equation modeling: Concepts, issues, and applications* (S. 56–75). Thousand Oaks, CA: Sage Publications.
- Whalley, L. J. & Deary, I. J. (2001). Longitudinal cohort study of childhood IQ and survival up to age 76. *British Medical Journal*, *322*(7290), 819–822. doi:10.1136/bmj.322.7290.819
- Wiederhold, K. A. (1991). Der Übergang von der Grundschule zu den weiterführenden Schulen – ein Problembereich für Kinder, Eltern und Lehrer. *Der Mathematikunterricht*, *37*(3), 6–19. doi:10.1007/978-3-531-91095-6_22
- Wigfield, A. & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review*, *12*(3), 265–310. doi:10.1016/0273-2297(92)90011-P
- Wigfield, A. & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. doi:10.1006/ceps.1999.1015
- Wigfield, A. & Eccles, J. S. (2002). *Development of achievement motivation*. San Diego: Academic Press.
- Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbretton, A. J. A., Feedman-Doan, C. & Blumenfeld, P. (1997). Change in children's competence beliefs and subjective task values across elementary school years: A 3-year study. *Journal of Educational Psychology*, *89*(3), 451–469. doi:10.1037/0022-0663.89.3.451
- Wigfield, A. & Karpathian, M. (1991). Who am I and what can I do? Children's self-concepts and motivation in achievement situations. *Educational Psychologist*, *26*(3-4), 233–261. doi:10.1080/00461520.1991.9653134
- Wild, E., Hofer, M. & Pekrun, R. (2006). Psychologie des Lernens. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie* (S. 203–267). Weinheim: Beltz.

-
- Willingham, W., Pollack, J. & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement*, 39(1), 1–37. doi:10.1111/j.1745-3984.2002.tb01133.x
- Wilson, R. S. (1983). The Louisville twin study: Developmental synchronies in behavior. *Child Development*, 54(2), 298–316.
- Wolf, H. & Riemann, R. (2008). Gen-Umwelt-Interaktion. In F. Neyer & F. Spinath (Hrsg.), *Anlage und Umwelt: Neue Perspektiven der Verhaltensgenetik und Evolutionspsychologie* (S. 85–101). Stuttgart: Lucius & Lucius.
- Wood, E., Nosko, A., Desmarais, S., Ross, C. & Irvine, C. (2006). Online and traditional paper-and-pencil survey administration: Examining experimenter presence, sensitive material and long surveys. *Canadian Journal of Human Sexuality*, 15(3-4), 147–155.
- Young, S. E., Stallings, M. C., Corley, R. P., Krauter, K. S. & Hewitt, J. K. (2000). Genetic and environmental influences on behavioral disinhibition. *American Journal of Medical Genetics*, 96(5), 684–695. doi:10.1002/1096-8628(20001009)96:5<684::AID-AJMG16>3.0.CO;2-G
- Zimmerman, M. A., Copeland, L. A., Shope, J. T. & Dielman, T. E. (1997). A longitudinal study of self-esteem: Implications for adolescent development. *Journal of Youth & Adolescence*, 26(2), 117–141. doi:10.1023/A:1024596313925