

Max-Planck-Institut für Bildungsforschung

Dissertation

# Mathematische Schülerleistung: Struktur, Schulformunterschiede und Validität

zur Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

im Fach Psychologie

Mathematisch-Naturwissenschaftliche Fakultät II

der Humboldt-Universität zu Berlin

**Dipl. Psych. Martin Brunner**

geboren am 14. Januar 1976 in Dachau

Präsident der Humboldt-Universität zu Berlin in Vertretung Prof. Dr. Jans Jürgen Prömel

Dekan: Prof. Dr. Uwe Küchler der Mathematisch-Naturwissenschaftlichen Fakultät II

Gutachter: 1. Prof. Dr. Oliver Wilhelm, Humboldt-Universität zu Berlin

2. Prof. Dr. Jürgen Baumert, Max-Planck-Institut für Bildungsforschung, Berlin

3. Prof. Dr. Olaf Köller, Humboldt-Universität zu Berlin

eingereicht: 30.09.2005

Datum der mündlichen Prüfung: 08.02.2006



### Namensnennung-NichtKommerziell-KeineBearbeitung 2.0 Deutschland

Sie dürfen:

- den Inhalt vervielfältigen, verbreiten und öffentlich aufführen

Zu den folgenden Bedingungen:



**Namensnennung.** Sie müssen den Namen des Autors/Rechtsinhabers nennen.



**Keine kommerzielle Nutzung.** Dieser Inhalt darf nicht für kommerzielle Zwecke verwendet werden.



**Keine Bearbeitung.** Der Inhalt darf nicht bearbeitet oder in anderer Weise verändert werden.

- Im Falle einer Verbreitung müssen Sie anderen die Lizenzbedingungen, unter die dieser Inhalt fällt, mitteilen.
- Jede dieser Bedingungen kann nach schriftlicher Einwilligung des Rechtsinhabers aufgehoben werden.

**Die gesetzlichen Schranken des Urheberrechts bleiben hiervon unberührt.**

Das Commons Deed ist eine Zusammenfassung des Lizenzvertrags <http://creativecommons.org/licenses/by-nc-nd/2.0/de/legalcode> in allgemeinverständlicher Sprache.

Haftungsausschluss <http://creativecommons.org/licenses/disclaimer-popup?lang=de>

## Abstrakt

Im Rahmen von drei Teilstudien wurde mathematische Schülerleistung aus einer differentialpsychologischen Perspektive untersucht. Die hierfür verwendeten Daten stammten von 29.386 deutschen Neuntklässlern, die am *Programme for International Student Assessment* (PISA) im Jahr 2000 teilnahmen.

In Studie 1 wurden ausgehend von Strukturtheorien kognitiver Fähigkeiten verschiedene Strukturmodelle mathematischer Schülerleistung konfirmatorisch geprüft. So wurde mathematische Schülerleistung in Form eines Nested-Faktormodell als additive Funktion einer mathematikspezifischen Fähigkeit ( $M'$ ) und der allgemeinen kognitiven Fähigkeit ( $g$ ) spezifiziert. Dieses Modell wies einen besseren Modellfit auf als das in der psychologischen Forschung dominierende Standardmodell. Für Letzteres wurde angenommen, dass Maße mathematischer Schülerleistung nur von einer generellen mathematischen Fähigkeit ( $M$ ) beeinflusst werden.

In Studie 2 wurden Schulformunterschiede mit konfirmatorischen Mehrgruppen-Faktormodellen untersucht. Schulformspezifische Mittelwertunterschiede in  $M$  waren im Standardmodell wesentlich stärker ausgeprägt als bei  $M'$  im Nested-Faktormodell. Weiterhin wurde eine schulformspezifische Differenzierungshypothese für  $M'$  untersucht. Entgegen der Erwartung konnte diese nur sehr eingeschränkt von den Daten gestützt werden.

In Studie 3 wurde die Validität mathematischer Schülerleistung im Hinblick auf soziodemografische und motivationale Schülermerkmale sowie Schulnoten analysiert. Bei Verwendung des Nested-Faktormodells resultierte ein im Vergleich zum Standardmodell wesentlich differenzierteres Befundmuster. So waren Geschlechterunterschiede (zu Gunsten der Jungen) in  $M'$  im Nested-Faktormodell deutlich stärker ausgeprägt als bei  $M$  im Standardmodell.

Implikationen und Perspektiven der drei Teilstudien werden für die psychologische Forschung, die Lehr-Lernforschung, die Konzeption von Schülerleistungsstudien sowie für die pädagogische Praxis diskutiert.

### *Schlüßswörter*

Schulleistung, kognitive Fähigkeiten, Mathematik, konfirmatorische Faktorenanalyse, Mehrgruppenmodelle, Differenzierungshypothese kognitiver Fähigkeiten, Geschlechterunterschiede, Selbstkonzept, Interesse, Schulnoten, *Programme for International Student Assessment* (PISA)

## Abstract

Three studies investigated mathematics achievement from an individual differences perspective, using data from 29,386 German ninth graders who participated in the 2000 cycle of the OECD's Programme for International Student Assessment (PISA).

In study 1, different structural models of mathematics achievement were derived from structural theories of cognitive abilities, and tested empirically using confirmatory methods. In a nested-factor model, mathematics achievement was specified to be an additive function of specific mathematical ability ( $M'$ ) and general cognitive ability ( $g$ ). This model provided a better fit than the standard model that predominates in psychological research, which assumes that measures of mathematical achievement are only influenced by general mathematical ability ( $M$ ).

In study 2, differences between types of schools were analyzed using confirmatory multigroup factor analytic models. Mean differences in  $M$  in the standard model were much stronger than in  $M'$  in the nested-factor model. A school-type-specific differentiation hypothesis for  $M'$  was also investigated. Contrary to predictions, the data provided only limited support for this hypothesis.

Study 3 analyzed the validity of mathematics achievement with respect to sociodemographic and motivational student characteristics and school grades. The nested-factor model yielded a much more differentiated pattern of results than the standard model. For example, gender differences (in favor of boys) were much more pronounced in  $M'$  in the nested-factor model than in  $M$  in the standard model.

The implications and future perspectives of studies 1 to 3 are discussed with respect to psychological and educational research, design of large-scale achievement studies, and educational practice.

### *Key Words*

achievement, cognitive abilities, mathematics, confirmatory factor analysis, multigroup models, differentiation hypothesis of cognitive abilities, gender differences, self-concept, interest, school grades, Programme for International Student Assessment (PISA)

## Inhaltsverzeichnis

1	Mathematische Schülerleistung als psychologisches Konstrukt?.....	10
2	Mathematische Schülerleistung aus psychologischer Sicht.....	15
2.1	Curriculare Definition mathematischer Schülerleistung .....	15
2.2	Die Begriffe Fähigkeit, Wissen und Lernen.....	16
2.3	Entwicklung mathematischer Fähigkeiten .....	17
2.4	Informationsverarbeitung beim mathematischen Problemlösen.....	18
2.5	Mathematische Fähigkeiten .....	20
2.6	Fazit und Arbeitsdefinition mathematischer Schülerleistung.....	21
3	Mathematische Schülerleistungstests .....	24
3.1	Geschichtlicher Überblick .....	24
3.2	Grundbildung versus curriculare Verankerung .....	27
3.3	Der PISA-Mathematiktest .....	29
4	Mathematische Schülerleistung und kognitive Fähigkeiten.....	33
4.1	Schülerleistung versus kognitive Fähigkeiten?.....	33
4.2	Strukturmodelle und mathematische Fähigkeiten.....	37
4.3	Integration und Differenzierung kognitiver Fähigkeiten .....	46
4.3.1	Allgemeine „Mechanismen“.....	46
4.3.2	Differenzierungshypothese.....	51
5	Mathematische Schülerleistung und schulisches Lernen .....	55
5.1	Modelle schulischen Lernens .....	55
5.2	Schülermerkmale und schulisches Lernen .....	58
5.2.1	Kognitive Fähigkeiten, Wissen und Wissenserwerbsprozesse .....	58
5.2.2	Lernmotivation.....	61
5.3	Mathematikunterricht.....	64
5.4	Schulformen als differenzielle Entwicklungsmilieus .....	67
6	Zusammenfassung vor dem Hintergrund der Forschungsfragen.....	71
7	Studie 1: Struktur mathematischer Schülerleistung.....	75
7.1	Mathematische Fähigkeiten.....	75
7.1.1	Empirische Befundlage und zu prüfende Strukturmodelle .....	75
7.1.2	Methode .....	80
7.1.3	Ergebnisse.....	85
7.1.4	Diskussion.....	87
7.2	Mathematische Fähigkeiten und (nicht mathematische) kognitive Fähigkeiten.....	90
7.2.1	Empirische Befundlage und zu prüfende Strukturmodelle .....	90
7.2.2	Methode .....	96
7.2.3	Ergebnisse.....	99
7.2.4	Diskussion.....	103
7.3	Zusammenfassende Diskussion zur Struktur mathematischer Schülerleistung .....	105
8	Studie 2: Schulformunterschiede .....	107
8.1	Niveauunterschiede .....	107

8.1.1	Empirische Befundlage und Hypothesen.....	107
8.1.2	Methode .....	110
8.1.3	Ergebnisse.....	113
8.1.4	Diskussion.....	118
8.2	Differenzierung der mathematikspezifischen Fähigkeit.....	120
8.2.1	Empirische Befundlage und Hypothesen.....	120
8.2.2	Methode .....	122
8.2.3	Ergebnisse.....	126
8.2.4	Diskussion.....	132
8.3	Zusammenfassende Diskussion zu Schulformunterschieden .....	135
9	Studie 3: Validität .....	138
9.1	Geschlechterunterschiede.....	140
9.1.1	Empirische Befundlage und Hypothesen.....	140
9.1.2	Methode .....	142
9.1.3	Ergebnisse.....	144
9.1.4	Diskussion.....	146
9.2	Familiärer Hintergrund .....	147
9.2.1	Empirische Befundlage und Hypothesen.....	147
9.2.2	Methode .....	148
9.2.3	Ergebnisse.....	149
9.2.4	Diskussion.....	151
9.3	Selbstkonzept.....	152
9.3.1	Das Bezugsrahmenmodell von Marsh (1986).....	152
9.3.2	Methode .....	155
9.3.3	Ergebnisse.....	156
9.3.4	Diskussion.....	159
9.4	Interesse .....	160
9.4.1	Ein erweitertes Bezugsrahmenmodell.....	160
9.4.2	Methode .....	162
9.4.3	Ergebnisse.....	163
9.4.4	Diskussion.....	168
9.5	Schulnoten .....	169
9.5.1	Empirische Befundlage und Hypothesen.....	169
9.5.2	Methode .....	171
9.5.3	Ergebnisse.....	172
9.5.4	Diskussion.....	173
9.6	Zusammenfassende Diskussion zur Validität.....	175
10	Gesamtdiskussion .....	179
10.1	Zusammenfassung.....	179
10.2	Schülerleistung versus kognitive Fähigkeiten? Ein Fazit.....	182
10.3	Implikationen und Perspektiven.....	183
10.3.1	Psychologische Forschung .....	183
10.3.2	Schülerleistungsstudien und Lehr-Lern-Forschung.....	187
10.3.3	Pädagogische Praxis.....	190
10.3.4	Schluss .....	193

Anhang .....	194
Testdesign bei der PISA-2000-Studie .....	194
Stichprobenvergleich: PISA 2000 versus Stichprobe dieser Arbeit.....	195
Testheftspezifische Korrekturparameter der WLE-Scores .....	196
Fehlende Werte bei den Maßen der kognitiven Fähigkeiten .....	199
Schulformspezifische deskriptive Statistiken .....	200
Modellparameter der Mehrgruppenfaktorenanalysen.....	202
Fähigkeitsgruppenspezifische Varianz der Faktorscores der allgemeinen kognitiven Fähigkeit .....	204
Differenzierungshypothesen: Prüfung psychometrischer Gesichtspunkte.....	205
Interkorrelationen der Maße kognitiver Fähigkeiten bei Kontrolle der Schulformzugehörigkeit .....	211
Vergleich der Korrelationen zwischen Schülermerkmalen und kognitiven Fähigkeiten in Abhängigkeit verschiedener faktorieller Strukturmodelle.....	212
Interkorrelationen der untersuchten Schülervariablen .....	213
Regressionsmodelle zur Vorhersage von kognitiven Fähigkeiten und Schulnoten .....	213
Literatur .....	216

## Abbildungsverzeichnis

Abbildung 1:	Prozessschritte beim mathematischen Problemlösen (nach Mayer, 1992).....	19
Abbildung 2:	Beispiel für eine technische Aufgabe .....	30
Abbildung 3:	Beispiel für eine rechnerische Modellierungsaufgabe.....	31
Abbildung 4:	Beispiel für eine begriffliche Modellierungsaufgabe .....	31
Abbildung 5:	Spearmans Zwei-Faktoren Theorie.....	37
Abbildung 6:	Thurstones Primärfähigkeiten.....	38
Abbildung 7:	Vernons hierarchisches Modell der Intelligenz .....	40
Abbildung 8:	Modell fluider und kristalliner Fähigkeiten (Horn & Noll, 1997).....	41
Abbildung 9:	Carrolls (1993) Drei-Stratum-Theorie.....	42
Abbildung 10:	Das Berliner Intelligenzstrukturmodell (nach Brunner & Süß, 2005).....	43
Abbildung 11:	Schematische Darstellung von Faktorintegration und Faktordifferenzierung (in Anlehnung an Baltes u.a., 1978).....	50
Abbildung 12:	Carrolls Modell (1963) des schulischen Lernens in Anlehnung an Harnishfeger und Wiley (1977) und Gruehn (2000).....	56
Abbildung 13:	Analyse der differenziellen Validität der kognitiven Fähigkeiten mathematischer Schülerleistung.....	74
Abbildung 14:	Mathematische Fähigkeiten .....	79
Abbildung 15:	Infit-Werte aller 117 Mathematikaufgaben aus dem PISA-2000-Test .....	86
Abbildung 16:	Modelle zur Überprüfung des Zusammenhangs von mathematischen Fähigkeiten und (nicht genuin mathematischen) kognitiven Fähigkeiten .....	94
Abbildung 17:	Modell 7a – Varianzzusammensetzung der stoffgebietspezifischen WLE-Scores.....	102
Abbildung 18:	Modelle zur Analyse von Schulformunterschieden .....	112
Abbildung 19:	Effektgrößen der schulformspezifischen Unterschiede in den mathematischen Fähigkeiten .....	116
Abbildung 20:	Faktormodell und schematische Darstellung der Differenzierung der mathematikspezifischen Fähigkeit ( $M'$ ).....	121
Abbildung 21:	Verteilung der Schüler über die Fähigkeitsgruppen in Abhängigkeit von der Schulform .....	123
Abbildung 22:	Modell zur Prüfung der Differenzierungshypothesen.....	125
Abbildung 23:	Generelle Differenzierungshypothese der mathematikspezifischen Fähigkeit ( $M'$ ).....	128
Abbildung 24:	Differenzierung der mathematikspezifischen Fähigkeit in Abhängigkeit des Fähigkeitsniveaus und der Schulformzugehörigkeit .....	131
Abbildung 25:	Gegenüberstellung der aus dem Modell zur schulformspezifischen Differenzierung der mathematikspezifischen Fähigkeit geschätzten und der tatsächlichen Differenzierung in dieser Stichprobe .....	134
Abbildung 26:	Modellspezifikation zur Prüfung der Geschlechterunterschiede.....	143
Abbildung 27:	Bezugsrahmenmodell in Anlehnung an Marsh (1986, S. 134) .....	153
Abbildung 28:	Erwarteter Effekt für das Bezugsrahmenmodell.....	154
Abbildung 29:	Erweitertes Bezugsrahmenmodell in Anlehnung an Nagy u.a. (in Druck) .....	162
Abbildung 30:	Korrelationen (mit 95%-Konfidenzintervall) zwischen kognitiven Fähigkeiten und Schülermerkmalen .....	176
Abbildung 31:	Standardmodell: Regressionsanalytische Kontrolle für die fluide Fähigkeit .....	184
Abbildung 32:	Varianzen der Faktorscores für die allgemeine kognitive Fähigkeit in Abhängigkeit des Fähigkeitsniveaus (in der Analysestichprobe) .....	204



Abbildung 33:	Mittelwerte und Varianzen der Faktorscores für allgemeine kognitive Fähigkeit in Abhängigkeit des Fähigkeitsniveaus und der Schulform.....	205
Abbildung 34:	Mittlere Schätzfehlervarianz der stoffgebietspezifischen WLE-Scores in Abhängigkeit des Fähigkeitsniveaus (in der Analysestichprobe) .....	206
Abbildung 35:	Mittlere Schätzfehlervarianz der stoffgebietspezifischen WLE-Scores in Abhängigkeit des schulformspezifischen Fähigkeitsniveaus .....	207
Abbildung 36:	Varianzen der stoffgebietspezifischen WLE-Scores in Abhängigkeit des Fähigkeitsniveaus (in der Analysestichprobe).....	209
Abbildung 37:	Varianzen der stoffgebietspezifischen WLE-Scores in Abhängigkeit des schulformspezifischen Fähigkeitsniveaus.....	210

## Tabellenverzeichnis

Tabelle 1:	Beschreibung kognitiver Fähigkeiten nach Carroll (1993) .....	20
Tabelle 2:	Zentrale Unterschiede zwischen den Strukturtheorien kognitiver Fähigkeiten .....	46
Tabelle 3:	Schüleranzahl, Geschlechterverteilung und mittleres Alter in Abhängigkeit der Schulform .....	82
Tabelle 4:	Kreuzklassifikation aller Aufgaben des Mathematiktests aus PISA 2000 nach mathematischen Stoffgebieten und Arten mathematischen Arbeitens .....	83
Tabelle 5:	Anzahl bearbeiteter Mathematikaufgaben pro Schüler .....	83
Tabelle 6:	Globaler Modell-Fit: $\chi^2$ - <i>Goodness-of-Fit</i> -Test und Informationskriterien .....	86
Tabelle 7:	Modell 2: Latente Korrelationen der operativen mathematischen Fähigkeiten/ Arten mathematischen Arbeitens .....	87
Tabelle 8:	Modell 3: Latente Korrelationen der stoffgebietspezifischen Fähigkeiten .....	87
Tabelle 9:	Mediane der Faktorladungen von Schülerleistungstests in 30 Studien (nach Treumann, 1974, S. 351, Tab. IV) .....	91
Tabelle 10:	Fluide Fähigkeit: Vergleich der Modell-Fit-Indizes für ein eindimensionales Rasch-Modell (1p) und ein eindimensionales, zweiparametrisches Birnbaum-Modell (2p) .....	97
Tabelle 11:	Deskriptive Statistiken der WLE-Scores zur Messung kognitiver Fähigkeiten .....	99
Tabelle 12:	Globale Modell-Fit-Indizes .....	100
Tabelle 13:	Standardisierte Faktorladungen ( $\lambda$ ) der Modelle 4 bis 7b .....	101
Tabelle 14:	Schulformspezifische Mittelwertunterschiede der allgemeinen kognitiven Fähigkeit in Effektgrößen d .....	107
Tabelle 15:	Schulformspezifische Mittelwertunterschiede der generellen mathematischen Fähigkeit (indiziert durch die globale Leistung in mathematischen Schülerleistungstests) in Effektgrößen d .....	108
Tabelle 16:	Globale Modell-Fit-Indizes unter verschiedenen Invarianzbedingungen .....	114
Tabelle 17:	Veränderung der globalen Modell-Fit-Indizes bei Einführung der Parameter <i>constraints</i> zur Prüfung der Invarianz .....	114
Tabelle 18:	Schulformspezifische unstandardisierte latente Mittelwerte ( $\mu$ ), korrespondierende 95%-Konfidenzintervalle [95%-KI] .....	115
Tabelle 19:	Test auf schulformspezifische Niveauunterschiede der kognitiven Fähigkeit mathematischer Schülerleistung .....	117
Tabelle 20:	Schulformspezifische Varianz ( $\sigma^2$ ) der mathematikspezifischen Fähigkeit ( $M'$ ) und korrespondierendes 95%-Konfidenzintervall [95%-KI] .....	120
Tabelle 21:	Generelle Differenzierungshypothese der mathematikspezifischen Fähigkeit: Deskriptive Modell-Fit-Indizes für die Analytestichprobe .....	127
Tabelle 22:	Generelle Differenzierungshypothese: Deskriptive Modell-Fit-Indizes für die simulierten Daten .....	127
Tabelle 23:	Inferenzstatistische Prüfung der generellen Differenzierungshypothese .....	129
Tabelle 24:	Schulformspezifische Differenzierungshypothese: Deskriptive Modell-Fit-Indizes .....	130
Tabelle 25:	Inferenzstatistische Prüfung der schulformspezifischen Differenzierungshypothese .....	132
Tabelle 26:	Korrelationen ( $r$ ) zwischen kognitiven Fähigkeiten (KF) und Geschlecht sowie korrespondierende $\chi^2$ -Differenzentests ( $\Delta\chi^2$ ) zur Prüfung der Hypothesen (Hpt) zu Geschlechterunterschieden .....	145

Tabelle 27:	Korrelationen ( $r$ ) zwischen kognitiven Fähigkeiten (KF) und Indikatoren des familiären Hintergrunds sowie korrespondierende $\chi^2$ -Differenzentests ( $\Delta\chi^2$ ) zur Prüfung der Hypothesen (Hpt) .....	150
Tabelle 28:	Items zur Messung des mathematischen und verbalen Selbstkonzepts.....	156
Tabelle 29:	Globale Modell-Fit-Indizes der Strukturmodelle zur Analyse des Bezugsrahmenmodells.....	157
Tabelle 30:	Standardmodell: Latente Korrelationen der kognitiven Fähigkeiten mit den fachspezifischen Selbstkonzepten.....	157
Tabelle 31:	Nested-Faktormodell: Latente Korrelationen der kognitiven Fähigkeiten mit den fachspezifischen Selbstkonzepten.....	157
Tabelle 32:	Analysen des Bezugsrahmenmodells für das Standardmodell und des Nested-Faktormodells: standardisierte Modellparameter (in Klammern: Modellparameter bei Kontrolle der Schulformzugehörigkeit).....	158
Tabelle 33:	Items zur Messung des mathematischen Interesses und des Interesses am Lesen.....	163
Tabelle 34:	Globale Modell-Fit-Indizes der Strukturmodelle zur Analyse des erweiterten Bezugsrahmenmodells.....	164
Tabelle 35:	Standardmodell: Latente Korrelationen der kognitiven Fähigkeiten mit den fachspezifischen Selbstkonzepten und Interessen .....	165
Tabelle 36:	Nested-Faktormodell: Latente Korrelationen der kognitiven Fähigkeiten mit den fachspezifischen Selbstkonzepten.....	165
Tabelle 37:	Standardmodell zur Prüfung des erweiterten Bezugsrahmenmodells: Standardisierte Regressionsgewichte und erklärte Varianz ( $R^2$ ) .....	166
Tabelle 38:	Nested-Faktormodell zur Prüfung des erweiterten Bezugsrahmenmodells: Standardisierte Regressionsgewichte und erklärte Varianz ( $R^2$ ) .....	167
Tabelle 39:	Korrelationen ( $r$ ) zwischen kognitiven Fähigkeiten (KF) und Schulnoten sowie korrespondierende $\chi^2$ -Differenzentests ( $\Delta\chi^2$ ) zur Prüfung der Hypothesen (Hpt).....	172
Tabelle 40:	PISA-O: Testdesign für den internationalen Vergleich der PISA-2000-Studie .....	194
Tabelle 41:	PISA-E: Testdesign für den nationalen Bundesländervergleich der PISA-2000-Studie.....	194
Tabelle 42:	Analysen zur Vergleichbarkeit der Analytestichprobe mit der PISA-2000-Stichprobe.....	196
Tabelle 43:	Korrekturparameter der WLE-Scores in Abhängigkeit der internationalen Testhefte.....	197
Tabelle 44:	Durch Testhefteeffekte erklärte Varianz ( $\eta^2$ ) der mathematischen Schülerleistung.....	198
Tabelle 45:	Muster gültiger ( $x$ ) und fehlender Werte ( $0$ ) der WLE-Scores der kognitiven Fähigkeiten für die Gesamtstichprobe und in Abhängigkeit der Schulform.....	199
Tabelle 46:	Deskriptive Statistiken getrennt für die Schulformen (SF) Hauptschule (HS), Realschule (RS) und Gymnasium (GY) .....	200
Tabelle 47:	Skalarinvariante Modellvarianten: Unstandardisierte ( $\lambda^*$ ) Faktorladungen, unstandardisierte Intercepts ( $\mu$ ) und schulformspezifische Residualvarianzen ( $\theta^2$ ) der manifesten Variablen .....	202
Tabelle 48:	Skalarinvariante Modellvarianten: Schulformspezifische Statistiken der latenten Variablen.....	203
Tabelle 49:	Vergleich der deskriptiven Statistiken mit und ohne Kontrolle der Schulformzugehörigkeit (SF).....	211

Tabelle 50:	Vergleich der Korrelationen zwischen Schülermerkmalen und den kognitiven Fähigkeiten mathematischer Schülerleistung in Abhängigkeit verschiedener faktorieller Strukturmodelle .....	212
Tabelle 51:	Interkorrelationen der kognitiven Fähigkeiten, der Schulformzugehörigkeit und der untersuchten Schulervariablen für das Standardmodell und das Nested-Faktormodell.....	213
Tabelle 52:	Vorhersage der kognitiven Fähigkeiten mathematischer Schülerleistung durch die analysierten Schülermerkmale und die Schulformzugehörigkeit .....	214
Tabelle 53:	Vorhersage der Schulnoten durch die kognitiven Fähigkeiten mathematischer Schülerleistung, die analysierten Schülermerkmale und die Schulformzugehörigkeit .....	215

## Danksagung

Das Gelingen einer Dissertation ist von vielen Menschen abhängig. Ihnen möchte ich an dieser Stelle danken.

An erster Stelle danke ich meinen Eltern, meiner Schwester Andrea und meinen Freunden, dafür, dass sie für mich da waren.

Prof. Dr. Jürgen Baumert danke ich für seine Förderung während meiner Zeit als Promotionsstipendiat am Max-Planck-Institut für Bildungsforschung, für seine konstruktiven Vorschläge und nicht zuletzt für sein Vertrauen in mich, den wirklich großartigen PISA 2000 Datensatz analysieren zu können. Prof. Dr. Oliver Wilhelm danke ich für die vielseitige, intelligente, pragmatische und intensive Unterstützung in allen Phasen dieser Dissertation. Prof. Dr. Olaf Köller danke ich für das Interesse an meiner Arbeit und die Bereitschaft diese zu begutachten.

Diese Arbeit würde anders aussehen, hätte ich nicht Möglichkeit gehabt, mit extrem schlaun und hilfsbereiten Kollegen zusammenzuarbeiten. Für das Lesen von sehr vielen Seiten und die exzellente Rückmeldung danke ich Ulrich Trautwein, Oliver Lüdtke, Stefan Krauss, Mareike Kunter und Michael Becker. Oliver und Ulrich sei insbesondere dafür gedankt, dass sie der Arbeit den entscheidenden Spin in Richtung „mathematische Schülerleistung“ gaben und wirklich bis zum Schluss alles gelesen haben. Stefan danke ich neben einigen sehr klugen Gedanken vor allem für den Glauben an mich und das Thema der Arbeit. Und Mareike danke ich für sehr wertvolle konzeptionelle Ideen. Nicht zuletzt danke ich Gabriel Nagy (wie auch nochmals Oliver) für viele interessante methodische und Cordula Artelt sowie Michael Schneider für viele inhaltliche Diskussionen. Peter Wittek danke ich für seinen sehr großen Einsatz beim Korrekturlesen.

Dem ganzen COACTIV-Team (Stefan, Mareike, Katrin Löwen, Uta Klusmann, Yi-Miau Tsai, Manuela Meermann und Petra Spengemann) sei für den Rückhalt gedankt und dafür, dass es Spass macht, mit euch zusammenzuarbeiten. Allen unseren wissenschaftlichen Hilfskräften und hier vor allem Madeleine Goslar, Annika Seehausen, Olaf Menzel, Charly Rosenbach und Maren Leifke danke ich für ihr riesiges Engagement.

Darüberhinaus sei aber auch allen Schülerinnen und Schülern gedankt, die an PISA 2000 teilnahmen und Cordula Artelt und Petra Stanat, die diesen „Supertanker“ von Studie koordinierten.

Zusammengenommen wäre ohne all diese Menschen meine Dissertation nicht möglich gewesen. Dankeschön, thank you, merci und grazie mille.

# 1 Mathematische Schülerleistung als psychologisches Konstrukt?

Stellen wir uns vor, wir hätten kein mathematisches Wissen. Könnten wir dann noch unsere beruflichen Anforderungen bewältigen? Würden wir die Zeitung verstehen, in der tagtäglich Tabellen und Grafiken abgebildet sind? Könnten wir die Vorschläge zur Altersvorsorge unseres Finanzberaters bei der Bank nachvollziehen? Die Antwort auf diese drei Fragen ist wahrscheinlich bei den meisten von uns „Nein“: Mathematik ist essenziell für die Teilhabe am gesellschaftlichen Leben.

In der Schule „lernt man für das Leben“, denn mathematisches Wissen und mathematische Fähigkeiten erwerben Kinder und Jugendliche primär in der Schule (Geary, 1995; Köller & Baumert, 2002; siehe auch Abschnitt 2.3). Was sie dort lernen, wird üblicherweise aus didaktischer Perspektive festgelegt (Köller & Baumert, 2002):

Die Spezifizierung dessen, was genau unter Schulleistungen zu verstehen ist, erfolgt typischerweise nicht aus psychologischer, sondern aus didaktischer Perspektive. (...) Psychologische Theoriebildungen zu Schulleistungen beschränken sich üblicherweise auf individuelle, gesellschaftliche und institutionelle Determinanten von Schulleistungen, wobei stillschweigend übergangen wird, worum es sich im Sinne psychologischer Konstrukte konkret bei Schulleistungen handelt.“ (Köller & Baumert, 2002, S. 757)

Diese Feststellung von Köller und Baumert betrifft insbesondere auch mathematische Schülerleistung<sup>1</sup> und ist erstaunlich, denn Lernen und Wissenserwerb sind zentrale Themen psychologischer Forschung. Es erleichtert sicherlich den Transfer psychologischer Forschungserkenntnisse in die pädagogische Praxis, wenn mathematische Schülerleistung an die Welt psychologischer Konstrukte angebunden ist. Die zentralen *Ziele dieser Arbeit* sind daher, mathematische Schülerleistung als psychologisches Konstrukt zu spezifizieren *und* aus dieser Perspektive besser zu verstehen.

Eine Beobachtung, die wahrscheinlich jede Leserin und jeder Leser dieser Arbeit schon einmal gemacht haben ist, dass nicht alle Schüler gleich gut im Fach Mathematik sind: Schüler<sup>2</sup> unterscheiden sich hinsichtlich ihrer mathematischen Schülerleistung. Diese interindividuellen Unterschiede der mathematischen Schülerleistung stehen im Mittelpunkt der vorliegenden Arbeit. Um (den differenzialpsychologischen Aspekt) mathematischer Schülerleistung besser zu verstehen, ist es notwendig, das Zustandekommen der interindividuellen Unterschiede theoretisch zu erklären. Hierzu bieten sich die Strukturtheorien kognitiver Fähigkeiten an. Allerdings besteht auch hier noch Forschungsbedarf. So stellt Carroll – eine Autorität auf dem Gebiet der Strukturforschung kognitiver Fähigkeiten – Folgendes fest:

In offering the (...) account of a theory of cognitive abilities as it might apply to mathematical thinking, I must admit to a degree of hesitancy and uncertainty. Despite six or seven decades of work in the psychometric tradition, relations between factor-analytically derived abilities and actual performances in various real-world domains, such as mathematics, have remained unclear. There is no doubt that relations exist, but exactly what they are and how they operate is not known as well as might be desired. (Carroll, 1996, S. 21)

---

<sup>1</sup> Da Schulleistungen primär von Schülern erbracht werden, wird in Anlehnung an Weinert (2001) nachfolgend von Schülerleistungen gesprochen.

<sup>2</sup> Zur sprachlichen Vereinfachung wird in dieser Arbeit anstelle von Schülerinnen und Schülern nur von Schülern gesprochen.

Auch in der Strukturforschung kognitiver Fähigkeiten sind also noch einige Fragen zur „real-world domain“ Mathematik offen. Das Ziel, mathematische Schülerleistung als psychologisches Konstrukt zu spezifizieren und besser zu verstehen, ist also sowohl aus der Perspektive der pädagogischen Praxis relevant als auch aus differenzialpsychologischer Forschungsperspektive interessant.

Im Hinblick auf die beiden Hauptziele der Arbeit werden drei zentrale Forschungsfragen verfolgt. Die erste Forschungsfrage fokussiert dabei auf die Spezifikation mathematischer Schülerleistung als psychologisches Konstrukt. Differenzialpsychologische Fähigkeitskonstrukte werden in der Regel in Form von (faktoriellen) Strukturmodellen konzeptualisiert. Deshalb lautet die erste Frage:

- (1) *Struktur mathematischer Schülerleistung: Welche kognitiven Fähigkeiten erklären interindividuelle Unterschiede in der mathematischen Schülerleistung?*<sup>3</sup> Sind es Fähigkeiten, die spezifisch für die mathematischen Stoffgebiete sind, oder ist es eine generelle mathematische Fähigkeit? Welche Rolle spielt die allgemeine kognitive Fähigkeit, die oftmals mit dem Schlagwort „Intelligenz“ bezeichnet wird? Antworten auf diese Fragen verknüpfen mathematische Schülerleistung mit differenzialpsychologischen kognitiven Fähigkeitskonstrukten.

Die nächsten beiden Forschungsfragen helfen, mathematische Schülerleistung aus der Perspektive kognitiver Fähigkeiten besser zu verstehen. Dabei ist zu bedenken, dass mathematische Schülerleistung in erster Linie das Ergebnis schulischer Lernprozesse (vgl. Abschnitt 2.3; Geary, 1995; Köller & Baumert, 2002) ist. Vor dem Hintergrund, dass ein großer Teil der interindividuellen Leistungsunterschiede bei mathematischen Schülerleistungstests mit dem Besuch unterschiedlicher Schulformen einhergeht (Baumert, Trautwein & Artelt, 2003), beschäftigt sich die zweite Forschungsfrage mit Schulformunterschieden:

- (2) *Schulformunterschiede: Wie unterscheiden sich Schüler, die unterschiedliche Schulformen besuchen, hinsichtlich der kognitiven Fähigkeiten mathematischer Schülerleistung?* Bei welchen kognitiven Fähigkeiten, die mathematischer Schülerleistung zu Grunde liegen, unterscheiden sich die Schulformen im Mittel am stärksten voneinander? Erzielen Schüler mit gleicher allgemeiner kognitiver Fähigkeit, die aber unterschiedliche Schulformen besuchen, die gleiche mathematische Schülerleistung? Wie sieht es mit der Heterogenität der Schülerschaft hinsichtlich ihrer mathematischen Fähigkeiten aus? Nimmt diese mit zunehmenden Niveau der allgemeinen kognitiven Fähigkeit zu? Ist diese so genannte „Differenzierung“ (z.B. Reinert, Baltes & Schmidt, 1965) von der Schulform abhängig? Über diese inhaltlichen Fragen hinaus wird bei den Analysen zu Schulformunterschieden die zentrale Voraussetzung zur Prüfung dieser Fragen untersucht: Inwiefern ist es möglich, die kognitiven Fähigkeiten mathematischer Schülerleistung schulformübergreifend mit identischen Strukturmodellen zu konzeptionalisieren?

Zusammenfassend werden die Ergebnisse zur zweiten Forschungsfrage dann ein sehr umfassendes und methodisch fundiertes Bild über Schulformunterschiede der kognitiven Fähigkeiten mathematischer Schülerleistung geben.

---

<sup>3</sup> Nachfolgend wird synonym zu der Bezeichnung „kognitive Fähigkeiten, die *interindividuelle Unterschiede mathematischer Schülerleistung erklären*“ auch die Formulierung *kognitive Fähigkeiten mathematischer Schülerleistung* oder *kognitive Fähigkeiten, die mathematischer Schülerleistung zu Grunde liegen*, verwendet.

Zusätzlich zu Schulformunterschieden sollen die kognitiven Fähigkeiten mathematischer Schülerleistung auch mit Blick auf ihre Zusammenhänge mit mehreren Schülermerkmalen beleuchtet werden. Ziel dieser Analysen ist es, die kognitiven Fähigkeiten mathematischer Schülerleistung inhaltlich weiter zu unterfüttern. Konkret lautet die dritte Forschungsfrage:

- (3) *Validität*:<sup>4</sup> Wie kovariieren die kognitiven Fähigkeiten mathematischer Schülerleistung mit soziodemografischen und motivationalen Schulervariablen oder Schulnoten? Sind Jungen besser in Mathematik als Mädchen? Wie stark sind Unterschiede des familiären Hintergrunds mit Maßen mathematischer Schülerleistung assoziiert? Wie hängen die kognitiven Fähigkeiten mathematischer Schülerleistung mit dem Interesse am Lesen oder der Einschätzung der eigenen verbalen Begabung zusammen? Positiv, negativ oder gar nicht?

Die vorliegende Arbeit ist natürlich nicht die erste, die diese Fragen zur externen Validität mathematischer Schülerleistung untersucht. Doch sie gehört zu den ersten Arbeiten, die die Validität mathematischer Schülerleistung bei gleichzeitiger Berücksichtigung der zu Grunde liegenden kognitiven Fähigkeiten mathematischer Schülerleistung analysiert. Denn in nahezu allen bisherigen Validitätsstudien mathematischer Schülerleistung (für einige Ausnahmen siehe zusammenfassend Gustafsson & Snow, 1997) wurden die zu Grunde liegenden kognitiven Fähigkeiten *nicht* getrennt betrachtet. Wenn diese kognitiven Fähigkeiten unterschieden werden, kann möglicherweise ein genaueres Bild der Zusammenhänge zwischen mathematischer Schülerleistung und Schülermerkmalen gezeichnet werden, als dies bisher der Fall war.

Daher wird im Rahmen der dritten Forschungsfrage also einerseits versucht, die bisherigen Befunde zur externen Validität mathematischer Schülerleistung zu replizieren. Andererseits soll untersucht werden, ob – im Vergleich zum bisherigen Vorgehen – alternative Konzeptionen der kognitiven Fähigkeiten mathematischer Schülerleistung einen detaillierteren Blick auf die externe Validität dieser Fähigkeiten ermöglichen.

Um die drei Forschungsfragen zu beantworten, gliedert sich die vorliegende Arbeit in einen theoretischen und in einen empirischen Teil. Im theoretischen Teil werden Theorien und Befunde aus der Forschung zur Struktur kognitiver Fähigkeiten, aus der pädagogisch-psychologischen Forschung und aus der pädagogischen Lehr-Lern-Forschung vorgestellt.

In Kapitel 2 wird zunächst das Problemfeld abgesteckt: Es werden mathematikdidaktische Gesichtspunkte von Schülerleistung skizziert, einige Begriffe (Fähigkeit, Wissen und Lernen) definiert und psychologische Forschungszugänge zu mathematischer Schülerleistung (Entwicklung, Informationsverarbeitung und Problemlösen sowie kognitive Fähigkeiten) vorgestellt. Vor diesem Hintergrund wird begründet, weshalb mathematische Schülerleistung in dieser Arbeit aus der Perspektive kognitiver Fähigkeiten untersucht wird.

Zum besseren Verständnis dafür, wie mathematische Schülerleistung mit standardisierten Tests gemessen wird, befasst sich Kapitel 3 mit mathematischen Schülerleistungstests.

---

<sup>4</sup> Der Begriff „Validität“ (im Sinne der externen Validität) wird in dieser Arbeit weiter interpretiert, als es die „Standards“ (American Educational Research Association, 1999, S. 9–13) vorsehen. Darin bezieht sich der Begriff „externe Validität“ darauf, inwiefern die empirisch gefundenen Beziehungen zwischen externen Variablen und Testscores konsistent sind mit den erwarteten Beziehungen hinsichtlich der Konstrukte, die der Test messen soll. Hingegen wird in der vorliegenden Arbeit der Begriff „Validität“ auf die kognitiven Fähigkeiten mathematischer Schülerleistung selbst angewendet und nicht auf einen bestimmten Test bezogen.



Kapitel 4 beschäftigt sich mit kognitiven Fähigkeiten. In Abschnitt 4.1 wird das Verhältnis von Schülerleistungen und kognitiven Fähigkeiten geprüft. Im nächsten Abschnitt (4.2) werden Strukturmodelle kognitiver Fähigkeiten vorgestellt, die in dieser Arbeit herangezogen werden, um mathematische Schülerleistung als psychologisches Fähigkeitskonstrukt zu spezifizieren (vgl. Forschungsfrage 1). Anschließend werden in Abschnitt 4.3 Wirkmechanismen erörtert, die erklären, wie Heterogenität in kognitiven Fähigkeiten entstehen kann (vgl. Forschungsfrage 2). Dabei wird insbesondere die Rolle der Lernumwelt (Abschnitt 4.3.1) und die Heterogenität kognitiver Fähigkeiten in Abhängigkeit der allgemeinen kognitiven Fähigkeit beleuchtet (Abschnitt 4.3.2).

Theorien und Befunde der pädagogisch-psychologischen Lehr-Lern-Forschung werden in Kapitel 5 dargestellt. Diese dienen zur theoretischen Untermauerung der Hypothesen, die mit Blick auf alle drei Forschungsfragen abgeleitet werden. Im Rahmen des Carroll-Modells (Carroll, 1963, 1989) wird das Zusammenwirken von kognitiven und motivationalen Schülermerkmalen und Merkmalen des Unterrichts beim schulischen Lernen allgemein erörtert (Abschnitt 5.1). Davon ausgehend wird in Abschnitt 5.2 dargestellt, welche Rolle individuelle Schülermerkmale beim Lernen einnehmen. Hierzu werden in Abschnitt 5.2.1 Prozesse beschrieben, die dem Wissens- und Fertigkeitserwerb zu Grunde liegen, und es werden kognitive Fähigkeiten als maßgebliche Determinanten dieser Lernprozesse identifiziert. Die Modelle schulischen Lernens wie auch die Theorien zum Wissens- und Fertigkeitserwerb stützen die Annahme, dass interindividuelle Unterschiede mathematischer Schülerleistung durch interindividuelle Unterschiede in kognitiven Fähigkeiten erklärt werden können (vgl. Forschungsfrage 1). In Abschnitt 5.2.2 werden motivationale Aspekte des Lernens diskutiert. Damit können Hypothesen abgeleitet werden, wie die kognitiven Fähigkeiten mathematischer Schülerleistung mit diesen motivationalen Schülermerkmalen kovariieren (vgl. Forschungsfrage 3). Neben individuellen Schülermerkmalen ist eine Hauptdeterminante mathematischer Schülerleistung der Mathematikunterricht. In Abschnitt 5.3 werden leistungsförderliche Aspekte des Mathematikunterrichts identifiziert. Schulformspezifische Unterschiede des Mathematikunterrichts (Abschnitt 5.4) unterfüttern die Annahme, dass sich die kognitiven Fähigkeiten mathematischer Schülerleistung zwischen den Schulformen im Niveau und ihrer Heterogenität unterscheiden können (vgl. Forschungsfrage 2).

Vor dem erarbeiteten theoretischen Hintergrund werden in Kapitel 6 die Theorien und Befunde der vorangegangenen Kapitel hinsichtlich der Forschungsfragen dieser Arbeit nochmals knapp zusammengefasst.

Konkrete Hypothesen werden im empirischen Teil formuliert und im Rahmen von drei Teilstudien empirisch überprüft. Studie 1 (Kap. 7) beschäftigt sich mit der ersten Forschungsfrage zur Struktur mathematischer Schülerleistung, Studie 2 (Kap. 8) befasst sich mit der zweiten Forschungsfrage zu Schulformunterschieden, und Studie 3 (Kap. 9) behandelt die dritte Forschungsfrage zur Validität der kognitiven Fähigkeiten mathematischer Schülerleistung.

Zur Beantwortung der Forschungsfragen werden multivariate statistische Verfahren eingesetzt. Diese Verfahren werden detailliert dokumentiert. Damit werden zwei Ziele verfolgt:

- Erstens wird transparent, welche Analysen zu welchen Ergebnissen führen. Somit ist eine weitestgehend objektive Beurteilung der Schlussfolgerungen möglich, die auf diesen Ergebnissen basieren.

- Zweitens ermöglicht erst die genaue Dokumentation die konzeptionelle Replikation der Ergebnisse in zukünftigen Studien, die sich – wie diese Arbeit – an der Schnittstelle von differenzialpsychologischer und pädagogisch-psychologischer sowie der pädagogischen Lehr-Lern-Forschung befinden.

Die Arbeit schließt mit einer Gesamtdiskussion (Kap. 10), in der die zentralen Ergebnisse der vorliegenden Arbeit zusammengefasst und mögliche Implikationen der Ergebnisse sowie Forschungsperspektiven erörtert werden.

## 2 Mathematische Schülerleistung aus psychologischer Sicht

Die vorliegende Arbeit beschäftigt sich mit mathematischer Schülerleistung aus der Perspektive psychologischer Konstrukte. Da Schülerleistungen in der Regel nach didaktischen Gesichtspunkten spezifiziert werden (Köller & Baumert, 2002), werden zunächst zentrale didaktische Aspekte mathematischer Schülerleistung skizziert (Abschnitt 2.1). Davon ausgehend werden häufig gebrauchte Begriffe der Lehr-Lern-Forschung definiert (Abschnitt 2.2). Im Anschluss werden drei psychologische Zugänge zu mathematischer Schülerleistung dargestellt (für eine umfassende Übersicht über Forschungsperspektiven auf mathematisches Denken siehe De Corte, Greer & Verschaffel, 1996; Schoenfeld, 1992; Stern, 1997; Sternberg & Ben-Zeev, 1996): Entwicklung (Abschnitt 2.3), Informationsverarbeitung und Problemlösen (Abschnitt 2.4), kognitive Fähigkeiten (Abschnitt 2.5). Abschließend werden in Abschnitt 2.6 diese drei Zugänge im Hinblick auf ihre Tragfähigkeit zur Beantwortung der Forschungsfragen der vorliegenden Arbeit bewertet, und es wird eine Arbeitsdefinition mathematischer Schülerleistung gegeben.

### 2.1 Curriculare Definition mathematischer Schülerleistung

Welche Lerninhalte aus didaktischer Sicht unter mathematischen Schülerleistungen subsumiert werden, wird meist in Lehrplänen konkretisiert (Köller & Baumert, 2002). Interessant ist dabei auf den ersten Blick die große Vielzahl an Mathematiklehrplänen in Deutschland: So ergab eine Recherche in der Lehrplandatenbank der Kultusministerkonferenz (KMK, <http://db.kmk.org/lehrplan/> am 14.2.2005) allein für die 9. Jahrgangsstufe 114 Treffer. Diese Zahl verwundert jedoch auf den zweiten Blick nicht weiter, wenn man bedenkt, dass für jede Schulform in jedem Bundesland ein eigener Lehrplan entwickelt wird. Die Tatsache, dass 114 verschiedene Lehrpläne in der 9. Jahrgangsstufe existieren, lässt vermuten, dass es „die“ mathematische Schülerleistung nicht gibt: Was als mathematische Schülerleistung in Schulform X in Bundesland Y verstanden wird, könnte etwas völlig anderes sein, als in Schulform X in Bundesland Z.

Doch ist das wirklich so? Eine Lehrplananalyse von Kunter (2005, Abb. 1.1) zeigte, dass mathematisches Verständnis und mathematische Problemlösefähigkeit in Deutschland länder- und schulformübergreifend als Ziele in den Lehrplänen und Rahmenrichtlinien der 8. Jahrgangsstufe genannt werden. Generell stehen neben diesen kognitiv anspruchsvollen Tätigkeiten in den deutschen Lehrplänen auch Ziele, wie das Üben und Vertiefen der (Grund-)Rechenkenntnisse und das Üben und Sichern mathematischer Verfahren (Kunter, 2005). Ein weiteres gemeinsames zentrales Merkmal ist, dass diese operativen Fähigkeiten im Rahmen eines Lehrplans unterrichtet werden, der sich sehr stark an den mathematischen Stoffgebieten ausrichtet (Neubrand u.a., 2001). Die mathematischen Stoffgebiete können auf einer ersten Beschreibungsebene in die Domänen Arithmetik, Algebra, Geometrie und Stochastik eingeteilt werden (Jordan u.a., 2006). Die Vermittlung von mathematischem Verständnis und Problemlösefähigkeiten sowie mathematischer Verfahren in einem nach Stoffgebieten gegliederten Lehrplan scheint also ein allgemeines Beschreibungsmerkmal deutscher Mathematikcurricula zu sein.

Wie internationale Curriculumvergleiche im Rahmen der SIMS- und TIMS-Studien zeigen, ist diese Beschreibung von Mathematiklehrplänen keineswegs auf Deutschland beschränkt: Auswahl und Sequenzierung von Stoffen von der Grundschule bis zur Sekundarstufe I sind staatenübergreifend in hohem Maße universell standardisiert (Baumert, 2001; Baumert, Köller, Lehrke & Brockmann, 2000; Bos & Postlethwaite, 2001; Schmidt, McKnight, Valverde, Houang & Wiley, 1996). Die Vermittlung von mathematischem Verständnis und Problemlösefähigkeiten scheint (in

westlichen Industrienationen) dabei eines der primären Anliegen des Mathematikunterrichts zu sein (Hiebert & Carpenter, 1992; Schoenfeld, 1992). Zudem erfolgt die Gliederung der Lerninhalte auch in anderen Nationen entlang mathematischer Stoffgebiete (z.B. National Council of Teachers of Mathematics, 2000).

Zusammenfassend kann man festhalten, dass aus didaktischer Perspektive mathematisches Verständnis und Problemlösefähigkeiten bundesländer- und nationenübergreifend weitestgehend übereinstimmend zur mathematischen Schülerleistung gezählt werden. Verständnis und Problemlösefähigkeiten basieren dabei auf Wissensinhalten, die im Rahmen eines nach mathematischen Stoffgebieten gegliederten Curriculums gelernt werden.

In den didaktischen Arbeiten zu mathematischer Schülerleistung werden häufig die Begriffe Fähigkeit und Wissen gebraucht. Diese Begriffe nehmen auch eine zentrale Stellung im Rahmen psychologischer Forschung ein. Der nächste Abschnitt hat daher das Ziel, Arbeitsdefinitionen dieser Begriffe aus einer psychologischen Perspektive zu geben.

## 2.2 Die Begriffe Fähigkeit, Wissen und Lernen

Was versteht man unter Fähigkeit, Wissen und Lernen?<sup>5</sup> Eine sehr brauchbare Definition von *Fähigkeit* liefert Carroll (1993, S. 16). Für Carroll stellt die Fähigkeit einer Person das gegenwärtige Potenzial dar, eine klar definierte Klasse von Aufgaben erfolgreich zu bearbeiten. Personen können sich dabei hinsichtlich ihrer Fähigkeit unterscheiden. Übertragen auf Maße mathematischer Schülerleistung bedeutet das: Wenn Schüler beispielsweise eine Vielzahl mathematischer Probleme erfolgreich lösen können, besitzen sie die Fähigkeit zum mathematischen Problemlösen.

Kognitive Fähigkeiten beziehen sich nach Carroll auf Aufgaben, für deren Lösung (mentale) Informationsverarbeitungsprozesse erfolgreich ausgeführt werden müssen. Zentrale Merkmale kognitiver Fähigkeiten sind die zeitliche Stabilität, der Grad der Generalität und der Grad der Universalität (Jäger, Süß & Beauducel, 1997): Zeitliche Stabilität bezieht sich darauf, inwiefern sich die Fähigkeit *nicht* über die Zeit verändert. Generalität ist ein Maß dafür, wie unterschiedlich die Aufgabenanforderungen sein können, für die eine Fähigkeit das Potenzial für eine erfolgreiche Bearbeitung hat. Universalität betrifft, ob eine Fähigkeit für alle Personen oder nur eine Teilmenge von Personen zur erfolgreichen Bearbeitung beiträgt.

Die wohl bekannteste Differenzierung von Wissen ist die Unterscheidung von prozeduralem Wissen und deklarativem Wissen (z.B. Anderson, 2001). *Prozedurales Wissen* ist das Wissen über die Art und Weise, wie man kognitive Aktivitäten und Operationen (z.B. Addieren, Multiplizieren, Konstruktion eines rechtwinkligen Dreiecks mit dem Thaleskreis) tatsächlich ausführt. *Deklaratives Wissen* umfasst das Wissen von Fakten und Dingen (Anderson, 2001): Im Fach Mathematik wäre das zum Beispiel das Wissen, wie der Satz von Pythagoras ( $a^2 + b^2 = c^2$ ) formuliert ist. Darüber hinaus hat sich in der pädagogischen Psychologie inzwischen der Begriff des konzeptuellen Wissens etabliert (Rittle-Johnson & Alibali, 1999; Rittle-Johnson, Siegler & Alibali, 2001): *Konzeptuelles Wissen* zeichnet sich durch eine große Anzahl an Wissens-elementen,

---

<sup>5</sup> Eine differenzierte Auseinandersetzung mit dem Fähigkeitsbegriff findet sich in der Zusammenfassung zweier Symposien von Du Bois (1969) und Green (1974). Der Begriff „Wissen“ wird in den Arbeiten von Dochy und Alexander (1995) sowie von Süß (1996) und der Begriff „Lernen“ in der Arbeit von Greeno, Collins und Resnick (1996) ausführlich erörtert.

die reiche Vernetzung der Wissens-elemente und Verständnis dafür aus, weshalb die Wissens-elemente miteinander verbunden sind. Konzeptuelles Wissen ist flexibel nutzbar, wohlorganisiert, generalisier- und transferierbar. Ein Beispiel hierfür wäre das Wissen darüber, dass der Satz von Pythagoras mit dem Thaleskreis zusammenhängt und warum das so ist.

*Lernen* kann als aktiver Konstruktionsprozess angesehen werden, bei dem Wissensstrukturen verändert, erweitert, vernetzt, hierarchisch geordnet oder neu generiert werden. Entscheidend für den Lernerfolg ist, dass der Lernende die Lerninhalte aktiv mental verarbeitet. Hierbei sprechen Greeno, Collins und Resnick (1996) von „conceptual learning“, Weinert (1996) sowie Baumert und Kollegen (Baumert & Köller, 2000; Baumert u.a., 2004) verwenden den Begriff „verständnisvolles Lernen“. Lernen von Mathematik bedeutet also, dass Schüler prozedurales und deklaratives mathematisches Wissen erwerben und dieses Wissen (idealerweise) in einer geordneten Struktur vernetzen (vgl. konzeptuelles Wissen).

Wie kann man das Verhältnis von Wissen, Fähigkeit und Lernen beschreiben? Forscher aus der Kognitions- und Lernpsychologie betrachten den prozeduralen Aspekt des Wissens weitestgehend synonym zum Begriff Fähigkeit: So sprechen Greeno, Smith und Moore (1993, S. 99) von „knowing“ als „ability to interact with things and other people in various ways“. In einer anderen vielzitierten Arbeit sehen Greeno, Collins und Resnick (1996) Wissen (im Sinne der Verfügbarkeit von Strukturen und Prozessen zur Informationsverarbeitung im Rahmen einer kognitiv-orientierten Lernforschung) als Grundlage dafür an, dass generelle Fähigkeiten wie Schlussfolgern, Problemlösen oder sprachliches Verständnis gezeigt werden können. Umgekehrt geht auch Carroll (1993, S. 510), als ein Vertreter des Fähigkeitskonzepts, davon aus, dass interindividuelle Unterschiede in Denkleistungen und beim Problemlösen (zumindest teilweise) abhängig sind von interindividuellen Unterschieden in der Verfügbarkeit und der Anwendung einer deklarativen und prozeduralen Wissensbasis der jeweiligen Inhaltsdomäne. Folgt man dem Argument einer synonymen Verwendung der Begriffe prozedurales Wissen und Fähigkeit, dann ermöglicht das Lernen von prozeduralem Wissen das Lernen von Fähigkeiten: „learning is improvement in that ability [to interact with things and other people in a situation]“ (Greeno u.a., 1993, S. 100). Auf Mathematik übertragen führt also Lernen zu einer Verbesserung der mathematischen Fähigkeit(en).

Mit Blick auf das Thema dieser Arbeit stellt sich nun die Frage, welche mathematischen Fähigkeiten und welches mathematische Wissen Schüler in der Schule lernen?

### **2.3 Entwicklung mathematischer Fähigkeiten**

Zur Beschreibung, welches mathematische Wissen und welche mathematischen Fähigkeiten in der Schule erworben werden, ist die Unterscheidung von Geary (1995, 1996, 2000) hilfreich: Geary unterscheidet (biologisch) *primäre* mathematische Fähigkeiten, die vornehmlich genetisch prädisponiert sind, und (biologisch) *sekundäre* mathematische Fähigkeiten, die in der Schule vermittelt und gelernt werden. Zur Beschreibung der Entwicklung mathematischer Schülerleistung können so im Wesentlichen zwei Zeitabschnitte unterschieden werden: vorschulische Entwicklung und die Entwicklung in der Grundschule und Sekundarschule.

*Vorschulische Entwicklung.* Zu den primären mathematischen Fähigkeiten (siehe auch Resnick, 1989) zählt Geary die Fähigkeit zur genauen Abschätzung von kleinen Mengen, ohne die Elemente zählen zu müssen (*numerosity*). Diese Fähigkeit ist Babys wohl angeboren, da sie bereits in

der ersten Lebenswoche zwischen „Mengen“ mit einem oder mehreren (bis zu vier) Elementen unterscheiden können. Bereits im Alter von fünf Monaten sind einige Kleinkinder sensitiv für Mengenzuwachs (Addition) und Mengenabnahme (Subtraktion) bei kleinen Mengen (mit bis zu drei oder vier Elementen). Geary bezeichnet dies als die primäre Fähigkeit zu *simple arithmetic*. Im Alter von ungefähr 18 Monaten können Kleinkinder ordinale Beziehungen (*ordinality*) verstehen, das heißt, sie haben Verständnis für Größer-kleiner-Relationen (für Quantitäten mit bis zu fünf Elementen) entwickelt. In diesem Alter sind Kinder auch in der Lage, drei bis vier Objekte zu zählen (*counting*), ohne dass sie sprechen können müssen.

Die primären mathematischen Fähigkeiten tragen im weiteren Entwicklungsverlauf zum Erwerb komplexerer Fertigkeiten im Bereich des Zählens, der Arithmetik und für den Aufbau des Zahlenverständnisses im Vorschulalter bei, sofern die Kinder die notwendigen Lernerfahrungen machen können. Wichtig ist hierbei, dass das zunächst sprachfreie Zahlensystem in die sich entwickelnde Sprachkompetenz integriert wird. Bereits vor Schuleintritt verfügen nach Geary (2000) Kinder über ein gutes Verständnis für das Zählen, für Ordinalität und für Kardinalität (das ist das Wissen, dass die letzte Zahl beim Abzählen von Objekten der Anzahl an Objekten in einer Menge entspricht).

Geary geht davon aus, dass die primären mathematischen Fähigkeiten biologisch-genetisch prädisponiert sind, auf hoch spezialisierten neurokognitiven Systemen (vgl. Gelman, 1990) basieren und das Ergebnis universeller, kulturübergreifender, evolutionärer Prozesse sind. Durch die Verarbeitung von quantitativen Informationen entwickeln sich die primären mathematischen Fähigkeiten. Viele dieser Informationsverarbeitungsprozesse sind spielerisch oder „passieren“ beim Spielen, machen Freude und motivieren die Kinder somit zur weiteren „Informationsverarbeitung“ (Geary, 1995).

*Entwicklung in der Grundschule und in der Sekundarstufe.* Die bis zum Schuleintritt entwickelten primären mathematischen Fähigkeiten können das Erlernen der sekundären mathematischen Fähigkeiten (hierzu gehören z.B. das Lösen von Problemen aus der Arithmetik, der Algebra oder der Geometrie) unterstützen (Geary, 1996). Zwei wesentliche Merkmale unterscheiden jedoch primäre von sekundären mathematischen Fähigkeiten (Geary, 1995). Einerseits vollzieht sich der Erwerb primärer mathematischer Fähigkeiten im Spiel und macht Spaß, wohingehend der Erwerb sekundärer mathematischer Fähigkeiten meist mit Anstrengung verbunden ist. Andererseits werden sekundäre mathematische Fähigkeiten (nahezu ausschließlich) durch schulische Lehr-Lern-Prozesse erworben, während die Entwicklung der primären mathematischen Fähigkeiten biologisch determiniert ist und kulturübergreifend gleichförmig verläuft (siehe auch Geary, 2000; Köller & Baumert, 2002).

Interessant ist die Frage, wie man diese in der Schule erworbenen sekundären mathematischen Fähigkeiten beschreiben kann (vgl. Mayer, 2004). Mit dieser Frage beschäftigt sich der nächste Abschnitt.

## **2.4 Informationsverarbeitung beim mathematischen Problemlösen**

Zur Beschreibung sekundärer mathematischer Fähigkeiten kann das Modell mathematischen Problemlösens von Mayer (1985, 1992, 1994; Mayer & Hegarty, 1996) herangezogen werden. Für das erfolgreiche Lösen eines mathematischen Problems sind nach Mayer einerseits verschiedene Wissensarten und andererseits mehrere kognitive Informationsverarbeitungsprozesse notwendig.

Die problemorientierte Anwendung der Wissensarten sowie die (erfolgreiche Ausführung der) Informationsverarbeitungsprozesse konstituieren die sekundären mathematischen Fähigkeiten sensu Geary.

Zu den Wissensarten gehört nach Mayer *linguistisches Wissen*, *semantisches Wissen* (das ist Alltagswissen), *schematisches Wissen* (als Wissen über mathematische Problemtypen, z.B. Flächenprobleme, siehe auch Mayer (1981), zu einer Abschätzung der Anzahl von Problemtypen), *strategisches Wissen* (als Techniken, die zum Planen und Überwachen des Lösungsprozesses eingesetzt werden, z.B. das Setzen von Zwischenzielen) und *prozedurales Wissen* (zum Ausführen mathematischer Operationen).

Für das Lösen eines mathematischen Problems sind zwei kognitive Prozesse zentral: Problemrepräsentation und Problemlösung (Abb. 1).

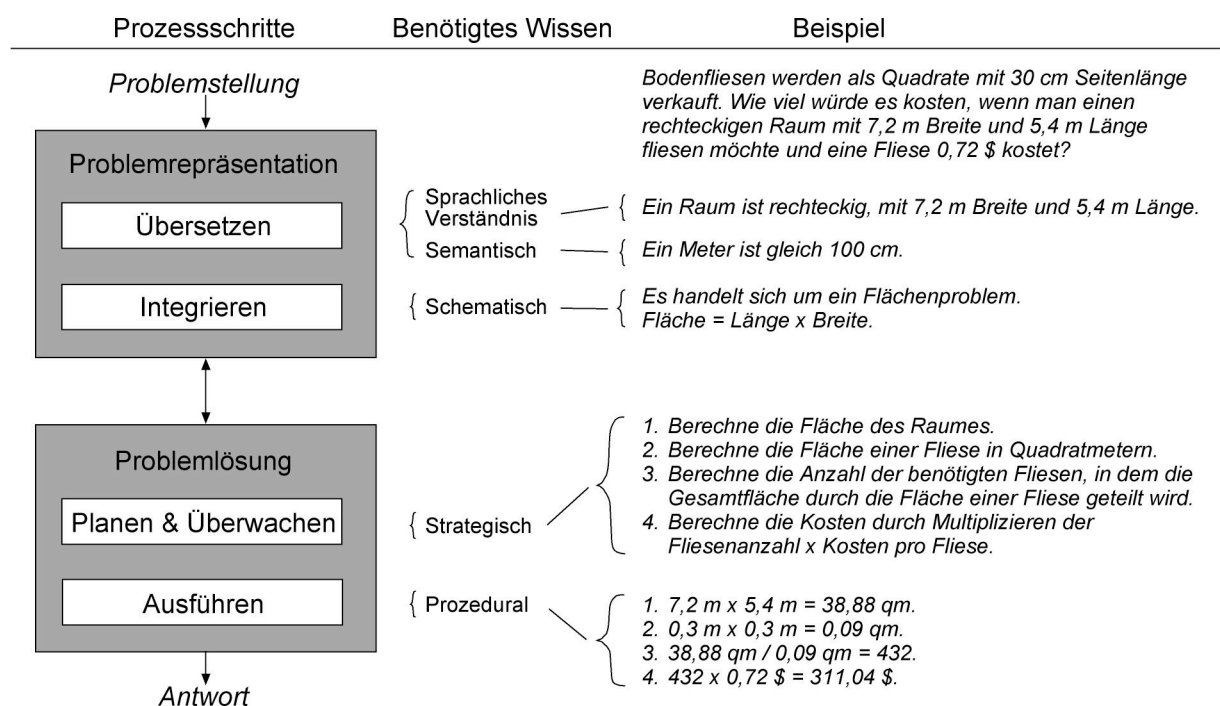


Abbildung 1: Prozessschritte beim mathematischen Problemlösen (nach Mayer, 1992)

Zur *Problemrepräsentation* muss man den Problemtext und gegebenenfalls vorhandene Abbildungen und Grafiken in eine mentale Repräsentation übersetzen. Der Aufbau der Problemrepräsentation vollzieht sich in zwei Teilschritten: Bei der *Problemübersetzung* konstruiert man mittels linguistischen und semantischen Wissens eine mentale Repräsentation. Bei der *Problemintegration* werden relevante Informationen aus der mentalen Repräsentation entnommen und mithilfe des schematischen Mathematikwissens und gegebenenfalls notwendiger zusätzlicher Schlussfolgerungen zu einer kohärenten mentalen Struktur organisiert. Bei der *Problemlösung* kommt man ausgehend von der Problemrepräsentation zur endgültigen Antwort. Zur Problemlösung sind zwei (zum Teil gleichzeitig ablaufende) Teilprozesse mit jeweils unterschiedlichen Wissensanforderungen notwendig. Beim *Planen* der Lösung und *Überwachen* der Planausführung greift man auf das strategische Wissen zurück und wendet dieses auf das konkrete Problem an: Es werden Problemlöseprozeduren ausgewählt und zusammengesetzt sowie mentale Ressourcen zugewiesen. Weiterhin

wird die eigene Problemlösetätigkeit überwacht und bewertet (siehe auch Mayer & Wittrock, 1996). Bei der *Ausführung* des Plans ist prozedurales Wissen in Form von mathematischen Algorithmen und Lösungsprozeduren notwendig.

Andere Modelle mathematischen Problemlösens unterscheiden sich in Detailfragen, stimmen aber in wesentlichen Bestimmungstücken mit dem Modell von Mayer überein. So wird auch bei den Modellen von Kintsch und Kollegen (Kintsch, 1994, 1998; Kintsch & Greeno, 1985; Nathan, Kintsch & Young, 1992) die besondere Rolle des Textverstehens und des Aufbaus eines mathematischen Problemmodells auf der Grundlage von schematischem mathematischem Wissen betont. Auch im modifizierten Modell „Mathematischen Modellierens“ von Schupp (1988), das die Autoren des PISA-Mathematiktests favorisieren (Blum u.a., 2004; Klieme, Neubrand & Lüdtke, 2001), wird dem Aufbau eines mathematischen Problemmodells besondere Bedeutung zugemessen. Darüber hinaus wird in diesem Modell wie auch im Modell von Mayer davon ausgegangen, dass die Anwendung von prozeduralem (und/oder konzeptuellem) Mathematikwissen und die Interpretation und Überprüfung der erarbeiteten Ergebnisse (dies korrespondiert mit den Überwachungsfunktionen) zentrale Prozessschritte beim mathematischen Problemlösen sind.

## 2.5 Mathematische Fähigkeiten

Während die Problemlöseforschung an den kognitiven Informationsverarbeitungsprozessen und dem hierfür notwendigen Wissen interessiert ist, fokussiert auf die Strukturforschung kognitiver Fähigkeiten die Produkte der Informationsverarbeitungsprozesse (vgl. z.B. Li & Schmiedek, 2001): zum Beispiel richtig oder falsch gelöste Mathematikaufgaben. Interindividuelle Unterschiede in der Problemlöseperformanz werden dadurch erklärt, dass sich Personen in einer oder mehreren kognitiven Fähigkeiten unterscheiden. Ein zentrales Kennzeichen dieser Forschung ist die Verwendung exploratorischer und konfirmatorischer Faktorenanalysen zur Ableitung und Prüfung von Strukturmodellen und der Konzeptualisierung kognitiver Fähigkeiten. In den korrespondierenden Modellen werden die kognitiven Fähigkeiten durch Faktoren (bzw. latente Variablen) repräsentiert (Carroll, 1993; Gustafsson & Undheim, 1996; Jensen, 1998).

Mit Blick auf Mathematikaufgaben geht zum Beispiel Carroll (1996) davon aus, dass je nach den Anforderungen einer Aufgabe unter anderem die Fähigkeiten Rechenfertigkeit oder quantitatives Reasoning lösungsrelevant sein können (siehe Tab. 1 für eine Beschreibung dieser Fähigkeiten): Entsprechend dieser Annahme erklären interindividuelle Unterschiede in diesen Fähigkeiten die beobachteten interindividuellen Unterschiede bei einer bestimmten Klasse von Mathematikaufgaben.

Tabelle 1: Beschreibung kognitiver Fähigkeiten nach Carroll (1993)

Fähigkeit	Beschreibung
Rechenfertigkeit	Fähigkeit im Umgang mit Zahlen. Hierzu gehört das Zählen von Objekten oder Erkennen von Zahlen, wie auch fortgeschrittene arithmetische Rechenoperationen (Addition, Subtraktion, Division, Multiplikation) mit einfachen Zahlen, Bruchzahlen oder Dezimalzahlen. Entscheidend bei diesen in der Regel eher einfachen Aufgaben sind Geschwindigkeit und Genauigkeit.
Quantitatives Reasoning	Schlussfolgerndes Denken mit Quantitäten oder mathematisch beschreibbaren Beziehungen. Notwendig ist hierfür die Anwendung von Konzepten aus der Algebra, Arithmetik oder Geometrie.



Diese Arbeit setzt sich mit mathematischer Schülerleistung aus der Perspektive kognitiver Fähigkeiten auseinander. Entsprechende Theorien werden deshalb ausführlich in Abschnitt 5.2 dargestellt. An dieser Stelle bleibt es zunächst bei der illustrativen Ausführung von Carroll.

## 2.6 Fazit und Arbeitsdefinition mathematischer Schülerleistung

In den vorangegangenen Abschnitten sind didaktische Aspekte mathematischer Schülerleistung skizziert, psychologische Begriffe definiert und drei psychologische Zugänge zu mathematischer Schülerleistung dargestellt worden. Nachfolgend werden die wesentlichen Unterschiede zwischen diesen drei Zugängen charakterisiert, um so ihre Tragfähigkeit zur Beantwortung der ersten Forschungsfrage in der vorliegenden Arbeit zu bewerten. Abschließend wird eine Arbeitsdefinition mathematischer Schülerleistung gegeben.

In entwicklungspsychologischen Untersuchungen mathematischer Schülerleistung sind Veränderungen über die Zeit der primäre Untersuchungsgegenstand. Von Interesse sind dabei Veränderungen mathematikspezifischer Informationsverarbeitungsprozesse, Veränderungen des mathematikspezifischen (deklarativen, prozeduralen, konzeptuellen) Wissens und natürlich auch Veränderungen der interindividuellen Unterschiede in der mathematischen Problemlöseleistung.

Im Unterschied hierzu spielen Veränderungen über die Zeit bei der Problemlöseforschung und der Strukturforschung kognitiver Fähigkeiten eine untergeordnete Rolle. Im Rahmen beider Forschungszugänge wird das Verhalten von Personen beim Lösen von Problemen und Aufgaben aus der Mathematik zu einem bestimmten Zeitpunkt untersucht, jedoch aus unterschiedlichen Blickwinkeln: Studien zum mathematischen Problemlösen adressieren mathematikspezifische Informationsverarbeitungsprozesse und Wissensarten (Mayer, 1985). Interindividuelle Unterschiede spielen eine untergeordnete Rolle. Wenn sie untersucht werden, werden sie durch interindividuelle Unterschiede in den postulierten Informationsverarbeitungsprozessen und Wissensarten erklärt (Mayer, 1985).

Im Rahmen der Strukturforschung kognitiver Fähigkeiten (z.B. Carroll, 1993) sind die beobachteten interindividuellen Unterschiede in der mathematischen Problemlöseleistung zentral und werden durch kognitive Fähigkeiten erklärt.

Ein mögliches Problem der Strukturforschung kognitiver Fähigkeiten beschreibt Mayer folgendermaßen:

The psychometric approach (...) defines mathematical ability as “what a math test measures” however, the psychometric definition is circular. It provides an excellent means for measuring mathematical ability, but it fails to provide an independent description of what is being measured. In contrast, the information-processing approach is based on task analysis. Any type of mathematics problem can be broken down into information-processing components (...) that are required for problem solution. (Mayer, 1985, S. 128)

Zwei Aspekte dieses Zitats von Mayer können jedoch kritisch betrachtet werden: Erstens erklären nicht alle postulierten Modelle der Strukturforschung kognitiver Fähigkeiten interindividuelle Unterschiede in der mathematischen Problemlöseleistung nur durch *eine* mathematische Fähigkeit (siehe hierzu Abschnitt 4.2).

Zweitens ist es notwendig, Informationsverarbeitungsprozesse und Wissensarten *reliabel* zu operationalisieren, um sie so einer Messung zugänglich zu machen. Nur so können interindividuelle Unterschiede in der mathematischen Problemlöseleistung durch interindividuelle Unterschiede in diesen Informationsverarbeitungsprozessen und Wissensarten erklärt werden. Allerdings steht hier die Problemlöseforschung selbst vor zwei – meiner Meinung nach – nicht gelösten Problemen. Diese werden nachfolgend stellvertretend am Modell von Mayer erörtert.

Es existiert eine Vielzahl von Wissensbegriffen, die mehr oder weniger geordnet und eher weniger aufeinander bezogen sind (für Übersichten siehe z.B. Dochy & Alexander, 1995; Süß, 1996): Gemeinhin werden die Begriffe deklaratives und prozedurales Wissen verwendet. In der pädagogisch-psychologischen Forschung kommt noch der Begriff des konzeptuellen Wissens hinzu (Rittle-Johnson & Alibali, 1999; Rittle-Johnson u.a., 2001). Eine empirisch validierte taxonomische Ordnung der Begriffe – wie sie im Rahmen der Strukturforschung kognitiver Fähigkeiten für eine Vielzahl von Fähigkeiten weitestgehend erreicht ist (Carroll, 1993; Flanagan, McGrew & Ortiz, 2000) – existiert in dieser Form im Bereich der Wissenspsychologie nicht. So differenziert zum Beispiel Mayer (1985) im Rahmen seines Modells mathematikunspezifische Wissensformen (linguistisches Wissen und semantisches Wissen) von mathematikspezifischen Wissensformen (schematisches, strategisches Wissen und prozedurales Wissen). Jedoch ordnet Mayer seine Wissensbegriffe *nicht* in die Kategorien deklaratives Wissen oder prozedurales Wissen ein.

Unabhängig davon ist es noch eine offene Frage, ob es gelingt, mit interindividuellen Unterschieden in Informationsverarbeitungsprozessen interindividuelle Unterschiede beim mathematischen Problemlösen zu erklären (Geary, 1994). Einerseits deshalb, weil es schwierig ist, Informationsverarbeitungsprozesse isoliert zu erfassen (siehe hierzu Lohman, 1994). Andererseits, weil es bisher noch sehr wenige Studien gibt, die explizit dieses versuchten (Geary, 1994). So ist mir keine Studie von Mayer bekannt, in der er die vier unterschiedenen Informationsverarbeitungsprozesse oder die fünf von ihm unterschiedenen Wissensarten *reliabel* und *eindeutig* erfasste und mit interindividuellen Unterschieden der mathematischen Problemlöseleistung in Beziehung setzte. Eine eindeutige Erfassung bedeutet hierbei auch, dass die verwendeten Maße von Maßen aus der Forschung kognitiver Fähigkeiten zu unterscheiden sind (siehe hierzu Abschnitt 4.1 und 10.3.1), denn dies ist der von Mayer formulierte Anspruch.

Mit Blick auf die Erklärung interindividueller Unterschiede der mathematischen Problemlöseleistung sind bisher also die empirischen Befunde hinter den theoretischen Erwartungen von Mayer zurückgeblieben. Der große Wert des Modells von Mayer für die vorliegende Arbeit liegt jedoch darin, dass es als heuristisches Modell gut geeignet ist, mathematisches Problemlösen zu beschreiben.

Zur empirischen Erklärung interindividueller Unterschiede der mathematischen Problemlöseleistung sind aber kognitive Fähigkeiten und Modelle aus der Strukturforschung kognitiver Fähigkeiten besser geeignet: Diese Modelle wurden auf Grundlage interindividueller Leistungsunterschiede bei kognitiven Aufgaben abgeleitet. Ihr Erklärungsabstand zur ersten Forschungsfrage dieser Arbeit ist also wesentlich kürzer als der von Modellen mathematisches Problemlösens. Die empirische Tragfähigkeit der spezifizierten Strukturmodelle zur Konzeptionalisierung der kognitiven Fähigkeiten mathematischer Schülerleistung wird in dieser Arbeit im Rahmen von allen drei Forschungsfragen geprüft.

Vor dem Hintergrund der gewählten Forschungsperspektive soll nun abschließend eine *Arbeitsdefinition* mathematischer Schülerleistung gegeben werden. Unter *mathematischer Schülerleistung* werden in dieser Arbeit die interindividuellen Unterschiede bei Aufgaben zur Messung mathematischer Schülerleistung verstanden. Die interindividuellen Unterschiede werden durch kognitive Fähigkeiten erklärt, wie sie in Strukturmodellen kognitiver Fähigkeiten konzeptualisiert werden. Da rivalisierende Strukturmodelle unterschiedliche Fähigkeiten zur Erklärung anbieten (vgl. Abschnitt 4.2), werden an dieser Stelle diese kognitiven Fähigkeiten zunächst nicht weiter spezifiziert.

### 3 Mathematische Schülerleistungstests

Wie wird die mathematische Schülerleistung eines Schülers bestimmt? Ein Indikator für die Schülerleistung sind natürlich die Noten, die auf schriftlichen oder mündlichen schulischen Leistungen basieren. Ein weiterer Indikator sind die Leistungen in standardisierten Schülerleistungstests. Ein zentrales Ziel beim Einsatz standardisierter Tests ist es, die Leistungen von Schülern vergleichbar zu machen, indem eine einheitliche Metrik verwendet wird.

Zum besseren Verständnis dafür, wie mathematische Schülerleistungen mit standardisierten Schülerleistungstests gemessen werden, befasst sich dieses Kapitel näher mit dieser Testform. Zunächst wird ein geschichtlicher Überblick über die Entwicklung von Schülerleistungstests gegeben (Abschnitt 3.1). Dabei wird näher auf internationale und nationale Vergleichsstudien von Schülerleistungen eingegangen und es werden in Abschnitt 3.2 Schwerpunkte bei der Testkonstruktion vorgestellt. Am PISA-Mathematiktest werden in Abschnitt 3.3 exemplarisch didaktische Gesichtspunkte zur Spezifikation von Schülerleistungen illustriert (vgl. Kap. 1; Köller & Baumert, 2002).

#### 3.1 Geschichtlicher Überblick

Nachfolgend wird stichpunktartig die geschichtliche Entwicklung von standardisierten Schülerleistungstests in den USA (dem „Mutterland“ standardisierter Tests) und in Deutschland nachgezeichnet. Dabei wird auch auf internationale Vergleichsstudien von Schülerleistungen eingegangen.

*Entwicklungen in den USA.* In den USA war ein erster Meilenstein in der Entwicklung standardisierter Schülerleistungstests die Einführung schriftlicher Examen an Schulen in Boston im Jahr 1845, die nach einheitlichen Korrekturanweisungen bewertet wurden (siehe z.B. Anastasi, 1968). Testtheoretische Grundlagen waren zu diesem Zeitpunkt allerdings noch nicht weit ausgereift. Erst 49 Jahre später im Jahr 1904 gab Edward L. Thorndike mit Herausgabe seines Buches *An Introduction to the Theory of Mental and Social Measurements* entscheidende methodische und statistische Impulse. Durch sein weiteres Wirken inspirierte er auch viele andere Forscher bei der Entwicklung standardisierter Tests. So konstruierte im Jahr 1908 Stone, ein Schüler Thorndikes, einen der ersten standardisierten Arithmetiktests (Ingenkamp, 1962), dem bald Tests für andere Fächer folgten (Anastasi, 1968).

Die Testentwicklung wurde in großem Maße durch die so genannte *survey-movement* motiviert. Ausgehend von Untersuchungen in der Stadt Pittsburgh im Jahr 1907 folgten immer mehr Studien, die Schulen derselben Stadt oder desselben Bundesstaates anhand von Testnormen verglichen, um so die Wirkung der öffentlichen Schularbeit zu untersuchen. Auch den diagnostisch tätigen Psychologen ermöglichten die Normdaten eine objektivierte Beurteilung einer Schülerleistung, denn ohne die Testnormen stand ihnen nur ihr eigener Erfahrungshintergrund zur Leistungseinschätzung zur Verfügung (Cronbach, 1984). Die *survey-movement* wurde dabei begleitet von der Einrichtung von pädagogischen Forschungsämtern, die ihrerseits die Entwicklung von Schülerleistungstests weiter vorantrieben (Ingenkamp, 1962). Zentrales Merkmal dieser Zeit war nach Cronbach (1984) die Dezentralisierung der Testentwicklung und Testanwendung. Jede Institution (Schulen, Universitäten usw.) entwickelte und administrierte eigene Tests oder konnte frei unter den verfügbaren Tests auswählen.

Jedoch zeichnete sich parallel auch eine langsame Entwicklung hin zu einer Zentralisierung der Testentwicklung und Testadministration ab (Cronbach, 1984). Im Jahr 1899 entstand das College Board (das damals noch College Entrance Examination Board hieß) als ein Zusammenschluss mehrerer Schulen (so genannter Preparatory Schools), die für das College vorbereiten. Das College Board wurde hauptsächlich deshalb gegründet, um auf mangelnde Übereinstimmung hinsichtlich der fachspezifischen Vorbereitung und unterschiedliche Bewertungsstandards zu reagieren. Auf Initiative des College Boards wurden im Jahr 1926 erstmals mathematische und verbale Aufgaben des SAT<sup>6</sup> administriert. Der SAT regelte seitdem neben dem *American College Test* (ACT) die Zulassung von Schülern zum College (Johnson, 1994). Im Jahr 1947 wurde der *Educational Testing Service* (ETS) gegründet (Anastasi, 1968), der seitdem für die Weiterentwicklung des SAT verantwortlich ist.

Standardisierte, zentral administrierte Schülerleistungstests haben in den USA also eine über 100-jährige Tradition und entscheidende Bedeutung für jeden Einzelnen, aber auch für die gesellschaftliche Entwicklung: Der SAT nimmt jedes Jahr maßgeblichen Einfluss auf die Bildungsbiografie und berufliche Karriere von Tausenden von Schülern, da das SAT-Ergebnis wichtig für die Zulassung zum College ist (Sternberg, 2003). *Large-Scale Assessment*-Programme, wie das vom US-Kongress initiierte *National Assessment of Educational Progress* (NAEP, NCES, 2005), informieren seit 1969 jährlich auf Grundlage repräsentativer Stichproben über den Bildungsstand der Schüler in verschiedenen Domänen (Lesen, Schreiben, Mathematik, Naturwissenschaften, Geschichte, Geografie) unterschiedlicher Jahrgangsstufen (4., 8. und 12. Klasse). Ergebnisse in solchen *Large-Scale Assessment*-Programmen dienen als Grundlage für bildungspolitische Entscheidungen.

*Entwicklungen in Deutschland.* Im Gegensatz zu den USA wurde standardisierten Schülerleistungstests in Deutschland zu Beginn des 20. Jahrhunderts nur wenig Aufmerksamkeit gewidmet (Ingenkamp, 1962). Zu den ersten deutschsprachigen Autoren, die Schülerleistungstests entwickelten, gehören Hylla und Bobertag (Bobertag & Hylla, 1928), die zum Beispiel mit ihren Tests T17 und T18 sprachliche und mathematische Schülerleistungen erfassten (Ingenkamp, 1962; Süllwold, 1983).

Einen ersten Aufschwung erfuhren Schülerleistungstests ab 1945 mit Ende des Zweiten Weltkriegs. Insbesondere die US-amerikanische Militärregierung unterstützte die Entwicklung und Anwendung von Schülerleistungstests: Es wurden Forschungsgelder bereitgestellt, Kontakte zwischen deutschen Forschungsinstituten und amerikanischen Experten im Rahmen eines „Sachverständigenprogramms“ hergestellt und die „Pädagogisch-psychologische Forschungsstelle Berlin“ errichtet. Daneben entstanden in München das „Testinstitut des Schulreferates der Stadt München“ und das „Deutsche Institut für Internationale Pädagogische Forschung“ (DIPF, ehemals „Hochschule für Internationale Pädagogische Forschung“) in Frankfurt am Main. In dieser Zeit wurden mehrere neue Schülerleistungstests entwickelt, amerikanische Verfahren ins Deutsche adaptiert oder bestehende Tests überarbeitet (z.B. die beiden oben genannten Tests T17 und T18). Mit Abbau der Dienststellen der US-Militärregierung wurden allerdings die Fonds aufgelöst, mit denen die Testentwicklung finanziert wurde. Dies führte zu einer Stagnation der Forschungsbemühungen fast aller Arbeitsgruppen. Ingenkamp zieht daher das Fazit, dass im Jahre 1962 nur noch an der Frankfurter Hochschule gute institutionelle Voraussetzungen für die Entwicklung von standardisierten Schülerleistungstests gegeben waren (Ingenkamp, 1962). Weitere Gründe für diese Stagnation waren methodische Mängel der Tests, aber vor allem die Test-

---

<sup>6</sup> SAT stand ehemals für Scholastic Assessment Test und später für Scholastic Aptitude Test. Mittlerweile ist die Abkürzung zum Eigennamen geworden und hat keine Bedeutung mehr (Sternberg, 2003).

skepsis von Lehrern, die Hand in Hand ging mit der fehlenden Vertrautheit von Lehrern im Umgang mit standardisierten Tests (Ingenkamp, 1968).

Gegen Ende der 1960er Jahre war die pädagogische Diagnostik jedoch wieder im Aufwind. Studien, die die Zuverlässigkeit des Lehrerurteils bei Übergangentscheidungen in Frage stellten, zeigten den Nutzen von Schülerleistungstests für diese diagnostischen Entscheidungen auf. Auch durch die Einführung der Gesamtschule entstand vermehrter Bedarf nach begleitender Evaluation mit standardisierten Testverfahren (Ingenkamp, 1968).<sup>7</sup> Neben dem DIPF befassten sich Ende der 1960er Jahre das Pädagogische Zentrum in Berlin wie auch das Max-Planck-Institut für Bildungsforschung in Berlin mit der Entwicklung von Schülerleistungstests (Ingenkamp, 1968).

Mittlerweile werden Schülerleistungstests in der pädagogischen Diagnostik verstärkt eingesetzt (z.B. zur Hilfe bei Schullaufbahnentscheidungen oder zur Lernstandsdiagnose), und es existieren mehrere veröffentlichte, standardisierte Testverfahren zur Messung mathematischer Schülerleistung (Süllwold, 1983). In der aktuellen Ausgabe des Standardwerks „Brickenkamp Handbuch psychologischer und pädagogischer Tests“ (Brähler, Holling, Leutner & Petermann, 2002) werden sechs Mehrfächertests und neun genuine Mathematik- und Rechentests aufgelistet.

Allerdings muss man dabei bedenken, dass standardisierte Schülerleistungstests *bei weitem* nicht in dem Maße eingesetzt werden, wie das in den USA der Fall ist. So ist zum Beispiel die Notwendigkeit einer Evaluation des deutschen Bildungswesens, die durch die Verwendung standardisierter Schülerleistungstests empirisch fundiert ist, sehr spät erkannt worden (Kahl & Spiewak, 2005). Zu großen Teilen machte erst das als schwach wahrgenommene Abschneiden der deutschen Schüler bei der *Third International Mathematics and Science Study* (TIMSS, z.B. Baumert, Bos & Lehmann, 2000) und vor allem beim *Programme for International Student Assessment* (PISA, z.B. Baumert, 2001) auf den Zustand des deutschen Bildungswesens aufmerksam. Eine Reaktion darauf war die Einführung von einheitlichen Bildungsstandards in Deutschland, die vom *Institut zur Qualitätsentwicklung im Bildungswesen* (IQB) wissenschaftlich begleitet wird (Ahnen, 2004).

*Internationale Vergleichsstudien von Schülerleistungen.* Die ersten Impulse für internationale Schülerleistungsvergleiche wurden in den 1950er Jahren des 20. Jahrhunderts gegeben. Organisationen wie die UNESCO oder die OECD analysierten Merkmale der Bildungssysteme (z.B. Curricula, Lehr-Lern-Methoden) unterschiedlicher Länder und verglichen als Indikator der Produktivität der Bildungssysteme den Anteil verschiedener Bildungsabschlüsse. Jedoch ließen diese Vergleiche nur bedingt Schlüsse über die Produktivität der Systeme zu, da mit gleichen Abschlüssen sehr unterschiedliche Leistungsstände der Schüler verbunden sein konnten. Daher trafen sich im Jahr 1958 mehrere Bildungsforscher, um die Möglichkeiten kulturübergreifender Schülerleistungsvergleiche mit standardisierten Testverfahren zu diskutieren: In der so genannten *pilot-study* wurde die positive Erfahrung gemacht, dass eine standardisierte Erhebung methodisch und organisatorisch möglich ist und dass die Ergebnisse fruchtbar für die Bildungspolitik sein können (Bos & Postlethwaite, 2001). Im Jahr 1964 wurde die erste internationale Vergleichsstudie, die *First International Mathematics Study* (FIMS), unter der Leitung der *International Association for the Evaluation of Educational Achievement* (IEA) in zwölf Staaten durchgeführt (Baumert, 2001; Bos & Postlethwaite, 2001). Dieser Studie folgten unter Leitung der IEA eine Reihe weiterer internationaler Vergleichsstudien, die Schülerleistungen in den Fächern Mathematik und Natur-

---

<sup>7</sup> In der ehemaligen Deutschen Demokratischen Republik wurden bereits seit 1965 regelmäßig repräsentative Schülerleistungsmessungen und zahlreiche Einzelstudien mit standardisierten Testverfahren zu vielfältigen Fragen der Unterrichtsforschung durchgeführt (Döbert & Geißler, 2000).

wissenschaften fokussierten. Zu den Studien, die mathematische Schülerleistungen untersuchten (für eine Übersicht siehe z.B. Stanat & Lüdtke, in Druck), gehören die *Second International Mathematics Study* (SIMS, z.B. Robitaille & Garden, 1989), die *Classroom Environment Study* (z.B. Helmke, Schneider & Weinert, 1986) sowie TIMSS (z.B. Baumert, Bos u.a., 2000; Kunter, 2005) mit ihren Erweiterungen (TIMSS-Video, siehe z.B. Klieme, Schümer & Knoll, 2001) und Nachfolgestudien (TIMSS-Repeat im Jahr 1999 und TIMSS 2003).

Parallel und in Konkurrenz zu den Vergleichsstudien der IEA hat die OECD mit PISA ein eigenes Forschungsprogramm initiiert (Baumert, 2001). In dreijährigen Zyklen mit wechselnden Schwerpunkten werden die Bereiche Lesen, Naturwissenschaften und Mathematik untersucht: An der ersten Erhebung im Jahr 2000 (PISA 2000) mit dem Schwerpunkt Lesen nahmen etwa 180.000 Schüler aus 32 Staaten teil (Baumert, Stanat & Demmrich, 2001). An PISA 2003 mit dem Schwerpunkt Mathematik beteiligten sich etwa 250.000 Schüler aus 41 Staaten (Prenzel, Drechsel, Carstensen & Ramm, 2004). Die internationale Vergleichsstudie wurde in Deutschland bei beiden Erhebungszyklen durch einen Bundesländervergleich ergänzt (Baumert, Artelt, Carstensen, Sibberns & Stanat, 2002; Prenzel u.a., 2004).

### 3.2 Grundbildung versus curriculare Verankerung

Ein wichtiger Gesichtspunkt bei der Messung von Schülerleistungen ist die Definition der zu erfassenden Gegenstandsbereiche (Stanat & Lüdtke, in Druck): Hier verläuft eine Trennlinie zwischen grundbildungsorientierten Tests und Tests, die sich an curriculärer Validität orientieren.

In den meisten IEA-Studien wird versucht, Testaufgaben zu konstruieren, die ein staatenübergreifendes Kerncurriculum erfassen (transnationale curriculare Validität, Baumert, 2001; Baumert, Köller, Lehrke & Brockmann, 2000; Bos & Postlethwaite, 2001; Schmidt, McKnight, Valverde, Houang & Wiley, 1996): Zentrales Ziel bei der Testkonstruktion ist daher, die Übereinstimmung zwischen dem internationalen Kerncurriculum und den für die Testaufgaben notwendigen Stoff, bzw. den notwendigen kognitiven Anforderungen zu maximieren. Ausgangspunkt hierfür ist die Beobachtung bei internationalen Curriculumvergleichen im Rahmen der SIMS- und TIMS-Studien, dass die Auswahl und Sequenzierung von Stoffen von der Grundschule bis zur Sekundarstufe I in hohem Maße universell standardisiert ist (Baumert, 2001; Bos & Postlethwaite, 2001; Schmidt u.a., 1996).

Ein weiteres Ziel bei der Konstruktion des TIMSS-Tests ist, Gedanken einer Grundbildungskonzeption mit einzubeziehen. Mit dem Begriff *Grundbildung* werden Basiskompetenzen adressiert, die die Lebensbewältigung im jungen Erwachsenenalter, kontinuierliches Weiterlernen in der Lebensspanne und das Lösen von Problemen in authentischen Anwendungssituationen ermöglichen sollen (Baumert u.a., 2001). Der TIMSS-Mathematiktest stellt daher einen „Kompromiss zwischen Anwendungsorientierung und curriculärer Anbindung“ dar (vgl. Baumert u.a., 2001, S. 19).

Während bei den IEA-Studien curriculare Validität zumindest teilweise angestrebt wurde und wird, ist die Leitidee bei der Konzeption und Konstruktion von den PISA-Tests ausschließlich die Erfassung der Grundbildung von Schülern in den Bereichen Lesen, Mathematik und Naturwissenschaften. Die Frage, ob die in PISA eingesetzten Testaufgaben transnational curricular valide sind, ist damit sekundär: Die Erfassung der Basiskompetenzen und damit auch die Test-

konstruktion folgt normativ einem bildungstheoretischen und didaktischen Konzept (Baumert u.a., 2001).

Allerdings macht es weder konzeptionell noch empirisch Sinn, Tests in disjunkte Kategorien „curricular valide“ oder „grundbildungsorientiert“ einzuteilen. Konzeptionell deshalb, weil natürlich auch Lehrpläne auf Vorstellungen basieren, welches Wissen und welche Fähigkeiten notwendig sind, um Jugendlichen am Ende ihrer Schullaufbahn die Teilhabe am beruflichen und gesellschaftlichen Leben zu ermöglichen (Stanat & Lüdtke, in Druck). Empirisch deshalb, weil einige Studien zeigen, dass es keinen bedeutsamen Unterschied macht, ob mathematische Schülerleistungen auf curricular verankerten Tests basieren oder ob ein Grundbildungskonzept der Testkonstruktion zu Grunde liegt (zusammenfassend z.B. Stanat & Lüdtke, in Druck):

- Artelt und Kollegen (Artelt, Brunner, Schneider, Prenzel & Neubrand, 2003) reanalysierten die PISA-2000-Daten zum Bundesländervergleich. Hierzu berechneten sie die Testleistung in Lesen, Mathematik und Naturwissenschaften der Schüler nur auf Grundlage der Items, die von Lehrplanexperten des jeweiligen Bundeslandes als curricular valide eingeschätzt wurden. Die auf Bundesländerebene aggregierten Testleistungen verglichen sie mit den ursprünglichen Testleistungen in der PISA-Untersuchung. Weder für Mathematik noch Naturwissenschaften noch Lesen resultierten statistisch signifikante Unterschiede.
- Die mathematischen Schülerleistungen beim nationalen PISA-Mathematiktest, der stärker am deutschen Mathematikcurriculum orientiert ist, korrelierten mit den Mathematikleistungen beim grundbildungsorientierten internationalen Test (für eine nähere Beschreibung siehe Abschnitt 3.3) mit .91 (Klieme, Neubrand & Lüdtke, 2001). Dieses Ergebnis werten Klieme u.a. als Hinweis darauf, dass internationaler und nationaler Mathematiktest die gleiche Fähigkeitsdimension erfassen.
- Stanat und Lüdtke berichten, dass sich die auf Staatenebene aggregierten Mittelwerte im internationalen Vergleich von PISA 2000 nicht ändern, wenn nur die jeweils lehrplankonformen Items eines Staates zur Mittelwertberechnung herangezogen werden.
- Bei einer methodisch ähnlichen Untersuchung mit Daten aus der TIMS-Studie änderten sich die auf Staatenebene aggregierten Leistungsmittelwerte ebenfalls nur marginal (z.B. Mullis u.a., 1998).

Zusammenfassend kann damit festgehalten werden, dass bei der Testkonstruktion mathematische Schülerleistung zwei Konzeptualisierungen existieren: curriculare Validität und die Grundbildung der Schüler. Mehrere Untersuchungen zeigen aber, dass die tatsächlich erzielten mathematischen Schülerleistungen weitestgehend unabhängig von diesem konzeptionellen Aspekt sind. Dies bedeutet, dass eine Einteilung von Tests in disjunkte Kategorien „curricular valide“ oder „grundbildungsorientiert“ empirisch gesehen nicht haltbar ist.

Bei der Diskussion curricularer Validität versus Grundbildung wurde bisher die Frage ausgeklammert, was es konkret bedeutet, wenn mathematische Schülerleistungen aus didaktischer Perspektive spezifiziert werden (vgl. Köller & Baumert, 2002). Wie sieht so ein mathematischer Schülerleistungstest aus? Zur Beantwortung dieser Fragen wird im nächsten Abschnitt beispielhaft der PISA-Mathematiktest vorgestellt.



### 3.3 Der PISA-Mathematiktest

Der Mathematiktest der Jahre 2000 und 2003 der PISA-Untersuchungen in Deutschland besteht aus einem internationalen Test und einem speziell für Deutschland entwickelten nationalen Test.

*Internationaler Test.* Mathematische Grundbildung wird im internationalen PISA-Rahmenkonzept definiert als die

(...) Fähigkeit einer Person, die Rolle zu erkennen und zu verstehen, die Mathematik in der Welt spielt, fundierte mathematische Urteile abzugeben und sich auf eine Weise mit der Mathematik zu befassen, die den Anforderungen des gegenwärtigen und künftigen Lebens dieser Person als konstruktivem, engagiertem und reflektierendem Bürger entspricht (Deutsches PISA-Konsortium, 2000, S. 47).

Mit der Wahl des Begriffs mathematische Grundbildung wird betont, dass es nicht so sehr um die Erfassung von mathematischen Kenntnissen und Fähigkeiten geht, wie sie schulische Curricula vermitteln. Mathematische Grundbildung zielt vielmehr auf Fähigkeiten ab, die die funktionale Anwendung von mathematischen Kenntnissen in unterschiedlichen Kontexten sowie Reflexion und Einsicht erfordern. Damit einher geht insbesondere die Fähigkeit, mathematische Probleme zu stellen, zu formulieren und zu lösen.<sup>8</sup> Durch das Curriculum vermitteltes Wissen und Fertigkeiten sind dabei notwendig, aber nicht hinreichend (Deutsches PISA-Konsortium, 2000).

Diese Konzeption mathematischer Grundbildung geht auf den niederländisch-deutschen Mathematikdidaktiker Hans Freudenthal zurück (Blum u.a., 2004; Klieme, Neubrand u.a., 2001): Freudenthal (1977) betont, dass mathematische Begriffe und Strukturen entstanden sind, um als Werkzeuge Ordnung in die Welt realer Phänomene zu bringen. Der Begriff „Welt“ umfasst die natürliche, technische, soziale und kulturelle Umwelt, aber auch die Welt der Mathematik selbst (Blum u.a., 2004).

Mathematische Grundbildung und darauf abzielende Testaufgaben werden anhand von zwei Hauptaspekten beschrieben (Blum u.a., 2004; Deutsches PISA-Konsortium, 2000). Der erste zentrale Aspekt sind allgemeine *mathematische Leitideen*, die die phänomenologischen Wurzeln der Mathematik abbilden (Blum u.a., 2004). Zu den Leitideen gehören Quantität, Veränderung und Beziehungen, Raum und Form und Unsicherheit. Diese vier Leitideen sind nicht identisch mit den mathematischen Stoffgebieten (Arithmetik, Algebra, Geometrie und Stochastik), aber damit verwandt. Als zweiter Aspekt betreffen *mathematische Kompetenzen* die Fähigkeiten, die in der Grundbildungskonzeption angesprochen werden. Hierzu zählen die Fähigkeiten, mit mathematischen Begriffen zu denken und zu argumentieren, mathematisches Modellieren, Probleme zu stellen und zu lösen, die Nutzung mathematischer Darstellungen und Hilfsmittel und die Kommunikation mathematischer Sachverhalte.

Bei der Operationalisierung der für den internationalen Vergleich eingesetzten PISA-Mathematikitems werden die einzelnen mathematischen Kompetenzen nicht getrennt erfasst, sondern sie sind im Verbund lösungsrelevant. Es können jedoch drei Kompetenzcluster gebildet werden, die sich danach unterscheiden, auf welchem Anspruchsniveau die jeweiligen Kompetenzen benötigt werden (Blum u.a., 2004). Aufgaben des *Kompetenzclusters 1 (Reproduktion)* fokussieren

---

<sup>8</sup> Insbesondere für mathematische Problemlöseprozesse betrachtet die OECD motivationale und affektive Aspekte der Mathematik, wie mathematisches Selbstkonzept oder Interesse, als wichtige und hilfreiche Voraussetzung. Allerdings gehören diese Aspekte nicht zum festen Bestandteil der Grundbildungsdefinition (Deutsches PISA-Konsortium, 2000).

auf die Wiedergabe von Definitionen oder Faktenwissen und technische Fertigkeiten wie Berechnungen, die Anwendung von Standardalgorithmen oder Formeln. Kompetenzen werden hierbei auf niedrigem Niveau benötigt; es handelt sich um mathematische Standardtätigkeiten. Zur Lösung von Aufgaben der *Kompetenzklasse 2 (Verbindungen)* müssen Aufgabenelemente zueinander in Beziehung gesetzt und Informationen verknüpft werden. Einige Kompetenzen werden auf mittlerem Niveau lösungsrelevant. Aufgaben der *Kompetenzklasse 3 (Reflexion)* erfordern teilweise (hoch) anspruchsvolle komplexe Tätigkeiten, Verallgemeinerungen und Reflexion.

*Nationaler Test.* Mathematische Leitideen und Kompetenzen sind die dominierenden Beschreibungsmerkmale für die Mathematikaufgaben, die im internationalen Test eingesetzt wurden. Bei den meisten Aufgaben des internationalen Testteils stehen außermathematische Anwendungssituationen im Vordergrund. Insbesondere Letztere werden im deutschen Mathematikunterricht eher nicht fokussiert. Deshalb wurden bei PISA 2000 und PISA 2003 Spezifika des deutschen Mathematikunterrichts wie dessen curriculare Struktur und vor allem die Betonung von kalkülorientierten Aufgaben im nationalen Testteil berücksichtigt, um ein umfassenderes Bild über den Leistungsstand der deutschen Schüler zu gewinnen (Neubrand u.a., 2001).

Von zentraler Bedeutung für den Aufbau und die Beschreibung des nationalen PISA-Mathematiktests ist die Unterscheidung von drei Aufgabenarten (*Arten mathematischen Arbeitens*), wie sie die deutsche PISA-Expertengruppe Mathematik eingeführt hat (Neubrand u.a., 2001). *Technische Aufgaben* (Abb. 2) erfordern Fertigkeiten und sind daher so konzipiert, dass die Ansätze vorgegeben sind und dass sie mittels bekannter Prozeduren (z.B. Rechnen oder geometrisches Konstruieren) bearbeitet werden können. Die Schwierigkeit technischer Aufgaben kann dabei variieren, da diese ein Spektrum algebraischer Verfahren bis hin zum Auflösen eines Gleichungssystems abdecken (Blum u.a., 2004; Klieme, Neubrand u.a., 2001).

Ein Rechteck ist 4 cm lang und 3 cm breit.  
Wie groß ist sein Flächeninhalt?

- 12 cm<sup>2</sup>
- 7 cm
- 12 cm
- 14 cm
- 7 cm<sup>2</sup>

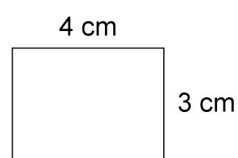


Abbildung 2: Beispiel für eine technische Aufgabe

Von technischen Aufgaben deutlich abzugrenzen sind rechnerische und begriffliche Modellierungsaufgaben. *Rechnerische Modellierungsaufgaben* (Abb. 3) sind Aufgaben, bei denen eine Mathematisierung notwendig ist. Hierzu müssen mathematische Lösungsansätze erstellt werden, die rechnerisch-prozedural aufzulösen sind. Beispiele hierfür sind Textaufgaben, deren Spektrum von eingekleideten Aufgaben bis zu komplexen Anwendungsproblemen reichen kann.

*Begriffliche Modellierungsaufgaben* (Abb. 4) benötigen für die Lösung das Herstellen von Zusammenhängen zwischen mathematischen Konzepten. Beim Verarbeitungsprozess dominieren qualitatives Denken und Schlussfolgern und nicht das Abarbeiten von Prozeduren (Klieme, Neubrand u.a., 2001, S. 146).

Hier siehst du ein Foto eines Bauernhauses mit pyramidenförmigem Dach. Nachfolgend siehst du eine Skizze mit den entsprechenden Maßen, die eine Schülerin vom Dach des Bauernhauses gezeichnet hat.



Der Dachboden, in der Skizze  $ABCD$ , ist ein Quadrat. Die Balken, die das Dach stützen, sind die Kanten eines Quaders (rechtwinkliges Prisma)  $EFGHKL MN$ .  $E$  ist die Mitte von  $\overline{AT}$ ,  $F$  ist die Mitte von  $\overline{BT}$ ,  $G$  ist die Mitte von  $\overline{CT}$  und  $H$  ist die Mitte von  $\overline{DT}$ . Jede Kante der Pyramide in der Skizze misst 12 m.

Berechne die Länge der Strecke  $\overline{EF}$ , einer der waagerechten Kanten des Quaders. Die Länge von  $\overline{EF} = \underline{\hspace{2cm}}$  m.

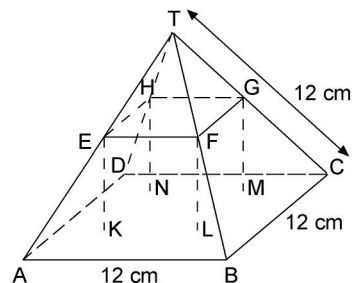


Abbildung 3: Beispiel für eine rechnerische Modellierungsaufgabe

Hier siehst du eine Karte der Antarktis.

Schätze die Fläche der Antarktis, indem du den Maßstab der Karte benutzt. Schreibe deine Rechnung auf und erkläre, wie du zu deiner Schätzung gekommen bist. (Du kannst in der Karte zeichnen, wenn dir das bei deiner Schätzung hilft.)



Abbildung 4: Beispiel für eine begriffliche Modellierungsaufgabe

Zur curricularen Strukturierung können die Mathematikaufgaben des nationalen und internationalen Testteils den *Stoffgebieten* der Sekundarstufe I zugewiesen werden. Im Einzelnen sind das Arithmetik, Algebra, Geometrie und Stochastik (Blum u.a., 2004; Jordan u.a., 2006; Klieme, Neubrand u.a., 2001).

Alle Aufgaben lassen sich zusätzlich auch nach der zur Lösung erforderlichen *curricularen Wissensstufe* beschreiben. Unter *Grundkenntnissen* verstehen Neubrand, Klieme und Lüdtke (2002) Grundrechenarten, einfachste geometrische Grundkenntnisse, die bereits in der Grundschule vermittelt werden oder aus dem Alltag bekannt sind. *Einfaches Wissen der Sekundarstufe I* umfasst Begriffe der Sekundarstufe I (z.B. der Bruchzahlbegriff), die in jeder Schulform zum Basiscurriculum gehören. Zum *anspruchsvollen Wissen der Sekundarstufe I* zählen fortgeschrittene Verfahren und Begriffe (im PISA-Mathematiktest waren das z.B. quadratische Gleichungen in der Algebra).

Zusammenfassend kann festgehalten werden, dass die internationale Rahmenkonzeption des PISA-Mathematiktests auf Fähigkeiten fokussiert, die die funktionale Anwendung von mathematischen Kenntnissen in unterschiedlichen Kontexten sowie Reflexion und Einsicht erfordern. Auch das nationale PISA-Mathematikrahmenkonzept zielt (mit Ausnahme der technischen Aufgaben) auf Problemlösefähigkeiten ab. Schwerpunkt des nationalen und internationalen PISA-Mathematiktests ist also nicht die Erfassung von im Curriculum vermittelten mathematischen Kenntnissen und Fähigkeiten, sondern die funktionale Anwendung (deklarativen und prozeduralen) Mathematikwissens zum Problemlösen.

## 4 Mathematische Schülerleistung und kognitive Fähigkeiten

Schüler unterscheiden sich in ihrer Performanz bei mathematischen Schülerleistungstests. Die Performanzunterschiede werden in dieser Arbeit aus der Perspektive kognitiver Fähigkeiten betrachtet. Die zentrale Annahme ist hierbei, dass interindividuelle Unterschiede in den kognitiven Fähigkeiten interindividuelle Unterschiede bei Maßen mathematischer Schülerleistung erklären. Zunächst beschäftigt sich daher Abschnitt 4.1 mit der Frage, worin sich Schülerleistungen von kognitiven Fähigkeiten unterscheiden. In Abschnitt 4.2 werden Strukturmodelle kognitiver Fähigkeiten vorgestellt und mathematische Fähigkeitskonstrukte in diesen Modellen beleuchtet (vgl. Forschungsfrage 1). Ziel von Abschnitt 4.3 ist es, zu klären, wie die in den Strukturmodellen beschriebene faktorielle Struktur zu Stande kommen kann. Dabei wird auch darauf eingegangen, wie sich die Heterogenität der kognitiven Fähigkeiten (in Form der Differenzierung) für Personen mit unterschiedlichem Niveau allgemeiner kognitiver Fähigkeit unterscheidet (vgl. Forschungsfrage 2).

### 4.1 Schülerleistung versus kognitive Fähigkeiten?

Mit der Fokussierung auf mathematische Problemlöseprozesse überlappt sich die internationale und nationale PISA-Rahmenkonzeption stark mit der Definition von Tests zur Messung von allgemeinen oder fluiden kognitiven Fähigkeiten, wie man sie in Intelligenztests findet: So erfordern beispielsweise nach dem renommierten Psychometriker Lloyd Bond (1989) diese Tests nur wenig deklaratives Wissen; die Schwierigkeit der Testitems resultiert aus der problembezogenen Anwendung des deklarativen Wissens. Damit in Einklang steht Carrolls (1993) Definition von Aufgaben zur Messung von quantitativem Reasoning. Lösungsrelevant sind für solche Aufgaben induktive oder deduktive Schlussfolgerungen und vor allem die problemorientierte Anwendung mathematischer Konzepte aus der Algebra, Arithmetik oder Geometrie.

Aus psychometrischer Perspektive können sich also Schülerleistungstests und kognitive Fähigkeitstests sehr ähnlich sein: Dies veranschaulichen Cooley und Lohnes (1976, S. 139) in ihrem Buch *Evaluation Research in Education* sehr deutlich. Sie legen ihren Lesern eine Tabelle mit Items vor, die aus einem Schülerleistungstest und einem Intelligenztest stammen. Der Leser soll nun die Quelle der Items angeben. Selbst Bond (1989) stellt hierzu fest, dass dies nahezu unmöglich ist.

Im Gegensatz hierzu werden in der pädagogisch-psychologischen Literatur häufig (z.B. Renkl & Stern, 1994; Spada & Wichmann, 1996; Weinert & Helmke, 1995a) allgemeine kognitive Fähigkeiten (Intelligenz) einerseits und Wissen oder Schülerleistungen andererseits unterschieden: Intelligenz und Wissen werden als Konstrukte konzeptuell klar voneinander getrennt (siehe auch Stern, 2001a, 2001b). Wenn die Konstrukte konzeptuell unterscheidbar sind, sollten auch ihre Operationalisierungen eindeutig zu diskriminieren sein. Vor dem Hintergrund der psychometrischen Ähnlichkeit und der gleichzeitigen konzeptuellen Unterscheidung ist es also eine wichtige Frage, welche operationalisierbaren Trennlinien zwischen Schülerleistungstests und Tests zur Messung kognitiver Fähigkeiten gezogen werden können (siehe hierzu auch die Zusammenfassung zweier Symposien, Du Bois, 1969; Green, 1974):

- (a) Eine *historische Trennlinie* in der Testkonstruktion zur Operationalisierung der Konstrukte existiert nicht (oder ist zumindest nicht markant ausgeprägt): Bei der Entwicklung der ersten Aufgaben ihres Intelligenztests forderten Binet und Simon zu Beginn des 20. Jahrhunderts Lehrer auf, die Arten von Aufgaben zu beschreiben, die für Schüler mit Lernschwierigkeiten

am schwierigsten zu lösen sind. Eine große Anzahl dieser Fragen und Aufgaben wurden in den ersten Intelligenztests aufgenommen. Um Schüler hinsichtlich ihrer Schülerleistung zu differenzieren, benötigt man natürlich ebenfalls Aufgaben, die für gute Schüler leicht und für schlechtere Schüler oder Schüler mit Lernschwierigkeiten schwierig zu lösen sind (Ceci, 1991). Daher sind viele Items aus Schülerleistungstests denen aus Intelligenztests sehr ähnlich, und Aufgaben im Stil von Binet und Simon sind auch heute noch in großer Zahl in modernen Intelligenztests enthalten (Humphreys, 1984, S. 223).<sup>9</sup> Zwei Argumente sprechen dafür, dass sich daran wenig geändert hat und ändern wird: Zum einen wollten andere Intelligenztestentwickler ihre Tests dadurch validieren, indem sie vergleichbare Resultate wie Binet und Simon bei der Vorhersage von Schulerfolg erzielen (Ceci, 1991). Zum anderen sind die Entwicklung von reliablen und validen Tests sowie die Testnormierung sehr kostspielig. Um Geld einzusparen und um das unternehmerische Risiko zu minimieren, werden daher oftmals die Testaufgaben von Intelligenztests in Schülerleistungstests (und umgekehrt) übernommen oder nur leicht adaptiert (vgl. Cronbach, 1984).

- (b) Eine Trennlinie hinsichtlich der verwendeten *psychometrischen Messmodelle* existiert nicht. Zur Messung kognitiver Fähigkeiten werden fast ausschließlich *Itemresponsemodelle* (IRT, z.B. Hambleton, Swaminathan & Rogers, 1991; van der Linden & Hambleton, 1997) und das Modell der klassischen Testtheorie (KTT, z.B. Lord & Novick, 1968) verwendet. Die Anwendung von IRT- und KTT-Modellen ist jedoch keineswegs auf die kognitive Fähigkeitsdomäne beschränkt: In *Large-Scale Assessment*-Programmen von Schülerleistungen wie PISA, TIMSS oder dem *National Assessment of Educational Progress* (NAEP) werden ebenfalls Verfahren aus der *Itemresponsetheorie* verwendet. Selbst in den Studien, in denen Intelligenz von Schülerleistungen strikt getrennt wird (z.B. Renkl & Stern, 1994; Weinert & Helmke, 1995a), werden die Testmodelle der klassischen Testtheorie herangezogen, um Schülerleistungen oder (Vor-)Wissen zu bestimmen. Unabhängig vom theoretischen Standpunkt zur Trennbarkeit der Konstrukte ist die zentrale statistische Annahme bei diesen Testmodellen, dass eine oder mehrere latente Variable(n) die Kovariation zwischen den beobachteten Testleistungen erklärt (McDonald, 1981, 1997, 1999). Im Rahmen von Theorien kognitiver Fähigkeiten wird die latente Variable dabei als kognitive Fähigkeit, in den *Large-Scale Assessment*-Programmen als Schülerleistung, in den zuletzt genannten Arbeiten als Vorwissen interpretiert.
- (c) Eine mögliche Trennlinie ist die *Neuartigkeit*. Aus psychometrischer Sicht kann dann keine Varianz auf interindividuelle Unterschiede im Vorwissen attribuiert werden, wenn alle Personen über das Wissen (inkl. das Wissen zur Verwendung der Antwortformate), das zusätzlich zu der kognitiven Fähigkeit für die Aufgabenlösung notwendig ist, in gleichem Ausmaß verfügen (Cattell, 1963): Dies kann bedeuten, dass im Extremfall alle Personen dieses Wissen besitzen oder niemand über dieses Wissen verfügt. Letzteres impliziert, dass die Aufgaben für alle Personen in gleicher Weise neuartig sind. Bei einer forschungspraktischen Umsetzung des Konzepts „Neuartigkeit“ stößt man jedoch auf zwei Probleme: Erstens ist es schwierig, alle Wissenskomponenten, die zusätzlich zu der kognitiven Fähigkeit lösungsrelevant sind, eindeutig a priori zu identifizieren. Zweitens sollten diese Wissenskomponenten auch empirisch nachweisbar sein, um überprüfen zu können, ob tatsächlich

---

<sup>9</sup> Hierbei ist jedoch einschränkend zu bemerken, dass es natürlich auch Aufgabentypen zur Erfassung mathematischer Fähigkeiten gibt (z.B. Zahlenreihen), die in dieser Form nicht oder nur sehr selten in mathematischen Schülerleistungstests zu finden sind. Die aufgabenspezifischen Varianzen werden im Faktormodell in den Residualtermen repräsentiert. Sofern also ein bestimmter Aufgabentyp in einer Aufgabenbatterie nicht überwiegt, kann davon ausgegangen werden, dass die verwendeten Aufgabentypen geringen Einfluss auf die Zusammensetzung der latenten Varianz nehmen.

deren Varianz konstant gehalten wird oder um statistisch für diese Varianzquellen zu kontrollieren. Allerdings gibt es keine Studien, die für die Neuartigkeit von Testaufgaben kontrollieren (siehe auch Stern, 2001b). So stellt Berg (2000) fest: „However, careful work that controls for the degree of task novelty for individuals across the life span has not been conducted.“ (Berg, 2000, S. 121)

Weiterhin wird die Aufgabenvertrautheit auch zur Unterscheidung von Aufgaben zur Messung von prozeduralem Wissen (korrespondierende Aufgaben sollten vertraut sein) und konzeptuellem Wissen (korrespondierende Aufgaben sollten neuartig sein) verwendet (Rittle-Johnson & Alibali, 1999; Rittle-Johnson, Sigler & Alibali, 2001). Unterscheidet man Tests alleine nach der Neuartigkeit, stellt sich somit die Frage, ob konzeptuelles Wissen und fluide Fähigkeiten identisch sind.

- (d) Eine weitere mögliche Trennlinie ist die *Breite der Wissensbasis* bzw. der *Erwerbskontext des Wissens*. Testleistungen spiegeln *immer* (auch) bisherige Lernerfahrungen wider (Anastasi, 1981; Carroll, 1993). Diese Lernerfahrungen können nach Anastasi genutzt werden, um die Breite der Wissensbasis und damit den Grad der Domänenspezifität zu definieren, die dem Testkonstrukteur zur Formulierung der Items zur Verfügung steht. Für die Konstruktion von Intelligenztests ist die Wissensbasis durch das gesamte Wissen definiert, das man durch außerschulisches *und* schulisches Lernen erwirbt (Anastasi, 1981; siehe auch Ceci, 1991). Hingegen sollten Schülerleistungstests Wissen und Fähigkeiten erfassen, die durch schulisches Lernen im Rahmen eines Curriculums erworben werden (vgl. Cleary, Humphreys, Kendrick & Wesman, 1975; Zieky, 1994). Um diese Idee tatsächlich in die eigene Forschung umzusetzen, stößt man jedoch wieder auf die gleichen Probleme wie beim Konzept der „Neuartigkeit“: Wie will man zum Beispiel bei 15-jährigen Jugendlichen für den Erwerbskontext von sprachlichen Fähigkeiten empirisch kontrollieren?
- (e) Cooley und Lohnes (1976) ziehen eine rein *funktionale Trennlinie* zwischen den verschiedenen Tests und damit auch den Konstrukten. Ein Test als Indikator für frühere Instruktion und Lernerfahrungen ist ein Schülerleistungstest. Falls dieser Test eingesetzt wird, um die aktuelle Kompetenz zu messen, ist es ein Fähigkeitstest. Wenn das Ziel die Vorhersage zukünftigen Verhaltens ist, ist es ein Befähigungstest. Bond hat diese Sichtweise pointiert so ausgedrückt: „Yesterday’s achievement is today’s ability and tomorrow’s aptitude.“ (1989, S. 429)
- (f) Eng assoziiert mit dieser funktionalen Unterscheidung ist die *regressionsanalytische Bedingung*, die Carroll (1974, 1993) nennt, um Befähigung und Schülerleistung zu unterscheiden: Ein Test misst dann die Befähigung eines Schülers, falls er bedeutsam zur Vorhersage in einem Schülerleistungstest beiträgt, wenn gleichzeitig für die Schülerleistung zu einem früheren Zeitpunkt kontrolliert wird. Befähigung ermöglicht also eine Vorhersage auf das (neu) Gelernte über das Vorwissen hinaus. Diese regressionsanalytische Unterscheidung liegt den Arbeiten von Renkl und Stern (1994) bzw. von Weinert und Helmke (1995a) zu Grunde.
- (g) Für Snow (1980) ist das *Wesen der Konstrukte* entscheidend: Warum kann man mit Schülerleistungstests oder Tests zur Messung kognitiver Fähigkeiten zukünftige Testleistungen vorher sagen? Nach Snow basiert die Performanz in Schülerleistungstests nicht nur auf einer mentalen Wissensstruktur, sondern auch auf mentalen Strukturierungs- und Organisationsprozessen, die wiederum den Erwerb neuen Wissens erleichtern. Dies erklärt nach Snow auch den Zusammenhang von kognitiven Fähigkeiten und Schülerleistung: Wenn Schülerleistungstests die funktionale Wissensanwendung oder Problemlösefähigkeit abbilden, erfor-

dern entsprechende Testaufgaben das Organisieren, Generalisieren, Adaptieren und die Anwendung von Wissen in neuen Kontexten. Diese kognitiven Funktionen sind aber gleichzeitig das, was nach Ansicht von Snow kognitive Fähigkeiten konstituiert (siehe auch Snow & Lohman, 1984). Snow plädiert deshalb für keine Trennung der Konstrukte, sondern vielmehr für: „We must stop acting as if cognitive abilities and achievements were arrays of separate mental garbage cans (...).“ (Snow, 1980, S. 44)

Glaser (1984), Carroll (1993, S. 510) sowie Greeno, Collins und Resnick (1996) vertreten einen ähnlichen Standpunkt: Für sie sind interindividuelle Unterschiede in Denkleistungen und beim Problemlösen (zumindest teilweise) abhängig von interindividuellen Unterschieden in der Verfügbarkeit und der Anwendung einer wohlorganisierten deklarativen, prozeduralen und konzeptuellen Wissensbasis der jeweiligen Inhaltsdomänen.

Was kann man nun als Fazit zu einer operationalisierbaren Trennbarkeit von mathematischen Schülerleistungen und kognitiven Fähigkeiten festhalten?

- (a) Mögliche Trennlinien von kognitiven Fähigkeiten und mathematischen Schülerleistungen sind die Neuartigkeit und die Breite der Wissensbasis. Jedoch ist der Grad der Neuartigkeit einer Testaufgabe kaum empirisch zu kontrollieren, wenn man davon ausgeht, dass trotz eines relativ homogenen Curriculums sich die Lernerfahrungen der Schüler (in Form des Mathematikunterrichts und der behandelten Mathematikaufgaben) zwischen Schulformen, aber auch zwischen Schulklassen innerhalb von Schulformen deutlich unterscheiden können (vgl. Abschnitt 5.3 und 5.4).

Weiterhin scheidet die Breite der Wissensbasis als Unterscheidungskriterium spätestens dann aus, wenn spezifischere Fähigkeiten und nicht generellere kognitive Fähigkeiten gemessen werden sollen: So ist die Beschreibung von quantitativem Reasoning nahezu identisch mit der Beschreibung mathematischer Grundbildung. Aus der Beschreibung der Konstrukte ist daher nicht klar, worin sich die korrespondierenden Operationalisierungen unterscheiden sollten.

- (b) Ihre Konstruktionsgeschichte oder die verwendeten Testmodelle (KTT oder IRT) sind genauso wenig wie eine funktionale Verwendung im Forschungskontext (vgl. Cooley & Lohnes, 1976) dazu geeignet, mathematische Schülerleistung von kognitiven Fähigkeiten zu differenzieren.
- (c) Der regressionsanalytische Ansatz von Carroll ist für längsschnittliche Fragestellungen relevant, kann aber nicht für querschnittliche Untersuchungen eingesetzt werden. Ein gewichtiges Argument gegen diesen Ansatz ist aber, dass Carroll keine Aussagen darüber macht, wie Testaufgaben beschaffen sein müssen, damit sie entweder Befähigung oder Schülerleistung erfassen.
- (d) Der Standpunkt von Snow, Glaser und Carroll, interindividuelle Unterschiede in Schülerleistungen bzw. Wissen (im weiteren Sinne) mit interindividuellen Unterschieden in kognitiven Fähigkeiten zusammenzubringen (und umgekehrt), stellt eine fruchtbare Synthese dar. In diesem Sinne spricht auch der renommierte Psychometriker Samuel Messick davon, dass

the coordinate measurement of developed cognitive abilities along with subject-matter achievement should contribute to the comprehensive understanding (...) of academic performance (Messick, 1984, S. 226).



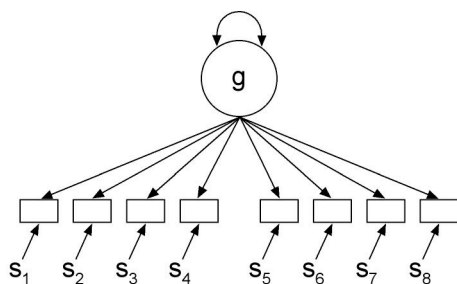
Die Perspektive dieses *Coordinate-Measurement*-Ansatzes motiviert, die empirisch fundierten Strukturmodelle kognitiver Fähigkeiten auf Maße mathematischer Schülerleistung anzuwenden. Diese theoretischen Modelle ermöglichen die Ableitung konkreter Hypothesen, welche kognitiven Fähigkeiten mathematischer Schülerleistung zu Grunde liegen. Eine Einbettung mathematischer Schülerleistung in Strukturmodelle kognitiver Fähigkeiten impliziert auch gleichzeitig eine Spezifikation mathematischer Schülerleistung in Termini psychologischer Konstrukte, wie sie Köller und Baumert (2002; vgl. Kap. 1) einfordern. Im nächsten Abschnitt werden deshalb entsprechende Strukturmodelle vorgestellt, und mathematische Schülerleistung wird aus dem Blickwinkel eines differenzialpsychologischen Fähigkeitskonstrukts betrachtet.

## 4.2 Strukturmodelle und mathematische Fähigkeiten

When one looks for a studied moment at the myriad of abilities that humans display, it's as if one were to look into the heavens on a clear night and become stirred by the ceaseless drift of the clouds of the Milky Way. On such a night one might be dimly aware that there is order and system in the celestial white. But where among the drifting haze might one draw dimensions to represent this order. At first there is no answer to this question, only befuddlement. The same is true for human abilities. They appear as free-floating swarms emerging from spaces of unknown many dimensions. (Horn, 1988, S. 645)

Das Hauptziel von Strukturmodellen kognitiver Fähigkeiten ist es, eine taxonomische Ordnung „in the celestial white“ kognitiver Fähigkeiten zu bringen (Gustafsson & Undheim, 1996), um so die Interkorrelationen zwischen Aufgaben zur Messung kognitiver Fähigkeiten möglichst sparsam zu erklären. Das Hauptaugenmerk der nachfolgend vorgestellten Strukturmodelle liegt darauf, anhand einer knappen Modellskizze den Status mathematischer Fähigkeiten im Rahmen dieser Modelle zu beleuchten. Hierzu werden die Modelle von Spearman und Thurstone vorgestellt, da sie grundlegend für später formulierte Modelle waren. Anhand der Modelle von Vernon (1964, 1965, 1969), von Cattell (1963) und Horn (1994; Horn & Noll, 1997), von Carroll (1993) und des Berliner Intelligenzstrukturmodells von Jäger (1967, 1982, 1984; Jäger, Süß & Beauducel, 1997) sollen differenzialpsychologische mathematische Fähigkeitskonstrukte betrachtet werden.

*Spearmans Zwei-Faktoren Theorie.* Einer der Gründungsväter der Forschung zur Struktur kognitiver Fähigkeiten ist Charles Spearman. Spearman (1904, 1927) nimmt an, dass eine allgemeine kognitive Fähigkeit existiert, die in unterschiedlichem Ausmaß an verschiedenen intellektuellen Aktivitäten beteiligt ist: Ein korrespondierender Faktor  $g$  erklärt die Varianzanteile, die allen Testaufgaben gemein ist. Ein weiterer so genannter Faktor  $s$  erklärt Varianz, die spezifisch für jede Aufgabe ist (Abb. 5). Spearman nennt daher sein Modell *Zwei-Faktoren Theorie* der Intelligenz.



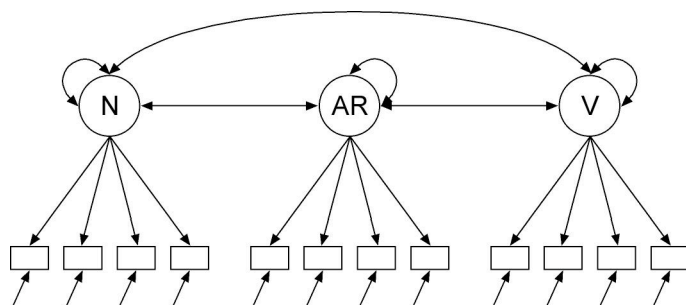
Allgemeine kognitive Fähigkeit ( $g$ ) und spezifische Fähigkeiten ( $s$ ).

Abbildung 5: Spearmans Zwei-Faktoren Theorie

Mit Spearman kann man also interindividuelle Unterschiede bei mathematischen Aufgaben durch  $g$  und durch mathematische Fähigkeiten erklären, die spezifisch für die eingesetzten Maße sind (dies entspricht den Faktoren  $s$  in Abb. 5). Weiterhin unterscheiden sich die eingesetzten Mathematikaufgaben darin, wie stark ihre Varianzen durch  $g$  oder die spezifischen Faktoren  $s$  beeinflusst werden und damit wie hoch sie auf den jeweiligen Faktoren laden. So lud in Spearmans Untersuchungen beispielsweise *problem arithmetic* hoch auf  $g$ , während „Simple addition“ niedrig auf  $g$  lud (Jensen, 1998, p. 35).

*Thurstones Primärfähigkeiten.* Die Annahme eines generellen Faktors  $g$  wurde in der Forschungswelt in der Folge stark kritisiert (Carroll, 1993). Thurstone (1936, 1938) schlug daher ein alternatives Strukturmodell vor, das mehrere spezialisierte Primärfähigkeiten (*primary mental abilities*) vorsieht. Jede dieser Primärfähigkeiten kann dabei die Interkorrelationen der Leistungen für eine bestimmte Aufgabenklasse erklären. Die von Thurstone identifizierten Primärfähigkeiten sind figural-räumliche Fähigkeiten (*spatial*), Wahrnehmungsgeschwindigkeit (*perceptual*), Rechenfertigkeit (*numerical*), Sprachliche Fähigkeiten (*verbal*), Gedächtnis (*memory*), Wortflüssigkeit (*word fluency*) und Induktives Reasoning (*inductive*). Auf einem weniger klar zu interpretierenden Faktor (Thurstone, 1938, S. 88) laden Aufgaben, die quantitative Reasoningprozesse erfordern (Gustafsson & Undheim, 1996, bezeichnen diesen Faktor als *Arithmetical Reasoning*).

In Thurstones Studien erklärten also (in Anlehnung an die Interpretation von Gustafsson und Undheim) zwei verschiedene mathematische Fähigkeiten (Rechenfertigkeit, quantitatives Reasoning mit Aufgaben aus der Arithmetik) interindividuelle Unterschiede bei Aufgaben zur Messung mathematischer Fähigkeiten (vgl. Abb. 6). Folgt man der Annahme, dass für eine bestimmte Klasse von Aufgaben kognitive Fähigkeiten unterschieden werden können, dann wäre es prinzipiell möglich, auch für weitere Klassen mathematischer Aufgaben (z.B. quantitatives Reasoning mit Aufgaben aus der Geometrie) spezifische mathematische Fähigkeiten zu identifizieren. Natürlich gilt diese Annahme auch für kognitive Fähigkeiten anderer Inhaltsdomänen.



N = Rechenfertigkeit, AR = quantitatives Reasoning mit Aufgaben aus der Arithmetik, V = verbale Fähigkeiten. Um die Abbildung übersichtlich zu gestalten, sind die weiteren von Thurstone identifizierten Primärfähigkeiten (siehe Text) nicht eingezeichnet.

Abbildung 6: Thurstones Primärfähigkeiten

Nicht zuletzt diese Annahme hat Forscher dazu motiviert, Studien zur Identifikation von Primärfähigkeiten durchzuführen. Wichtigstes Ergebnis dieser Untersuchungen ist, dass über 60 Primärfähigkeiten gefunden wurden (Carroll, 1993; Ekstrom, French & Harman, 1979). Mit Blick auf mathematische Fähigkeiten gehörten hierzu einerseits die Replikation des Faktors Rechenfertigkeit und andererseits der Faktor General Reasoning (Ekstrom u.a., 1979), der in erster Linie

quantitative Reasoningfähigkeiten misst (zusammenfassend siehe auch Geary, 1994). Die aufgabenspezifischen Fähigkeiten (siehe im Modell von Spearman) spielen in Thurstones Modell wie auch in den zeitlich später ausgearbeiteten Modellen keine entscheidende Rolle mehr.

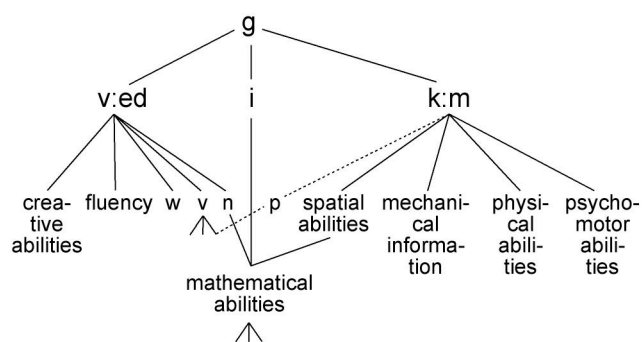
*Hierarchische Faktorenanalysen.* Nachfolgende Strukturmodelle haben versucht, die Interkorrelationen zwischen der anwachsenden Zahl an Primärfähigkeiten durch generellere Fähigkeitskonstrukte zu erklären. Hierzu wurden hierarchische Faktormodelle verwendet, die Gustafsson und Undheim (1996) in zwei Gruppen einteilen. Bei der ersten Analyseverfahren (ein *Bottom-up*-Ansatz) versucht man die Interkorrelationen der Faktoren erster Ordnung mit Faktoren zweiter Ordnung zu erklären. Sollten diese Faktoren zweiter Ordnung selbst wieder interkorreliert sein, wird eine dritte Hierarchieebene von Faktoren eingeführt werden, die die Interkorrelationen zwischen den Faktoren zweiter Ordnung erklärt und so fort. Diese Analyseverfahren wurde von den amerikanischen Forschern eingesetzt und hatte maßgeblichen Einfluss auf das Strukturmodell kristalliner und fluider Fähigkeiten (Cattell, 1963; Horn, 1988, 1994; Horn & Noll, 1997) wie auch der Entwicklung der Drei-Stratum-Theorie von Carroll (1993).

Die zweite Analyseverfahren (*hierarchische Gruppenfaktorentechnik*; ein *Top-down*-Ansatz) wurde von den britischen Forschern zur Untersuchung der Struktur kognitiver Fähigkeiten favorisiert. Bei dieser Methode werden so viele orthogonale Faktoren eingeführt, bis alle Interkorrelationen zwischen den manifesten Testaufgaben vollständig erklärt sind: Der erste Faktor erklärt dabei die meiste Varianz, ist also der breiteste und generellste Faktor. Für den Einfluss dieses ersten Faktors wird dann statistisch kontrolliert. Aus den verbleibenden Residualinterkorrelationen wird dann der zweite Faktor extrahiert. Die weiteren, zunehmend schmäleren Faktoren werden nach dem gleichen Prinzip gebildet.

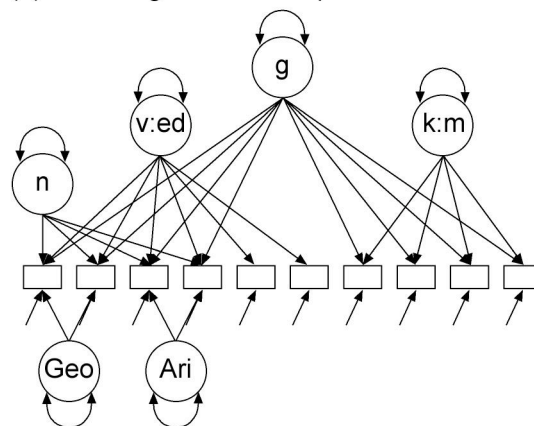
*Vernons hierarchisches Modell kognitiver Fähigkeiten.* Die einflussreichste Zusammenfassung der britischen Forschung zur Struktur kognitiver Fähigkeiten stammt von Philip E. Vernon (Gustafsson & Undheim, 1996). Auf Grundlage seiner Literaturreview und eigener Forschung hat Vernon (1964, 1965, 1969) ein hierarchisches Strukturmodell kognitiver Fähigkeiten entwickelt (Abb. 7a). An der Spitze des Modells steht die allgemeine kognitive Fähigkeit (*g*), die vornehmlich durch Reasoningaufgaben markiert wird. Im Modell sind auch zwei bedeutsame allgemeinere Fähigkeiten enthalten: verbal-numerische Fähigkeit (*v:ed*; *verbal-numerical-educational*) und praktisch-mechanisch-räumlich-physikalische Fähigkeit (*k:m*; *practical-mechanical-spatial-physical*). Wichtig ist hierbei, dass Vernon *v:ed* noch weiter differenziert und konzeptuell eine generelle mathematische Fähigkeit (*n*, die Vernon, 1969, S. 22, als *number* bezeichnet) im Modell vorsieht. Diese ist wiederum spezifischeren mathematischen Fähigkeiten (*mathematical abilities*, z.B. Geometrie, Algebra, Arithmetik, vgl. Vernon, 1964) hierarchisch übergeordnet.

Aus den theoretischen Darstellungen von Vernon ist es schwierig, eine *eindeutige* faktorielle Repräsentation von Aufgaben zur Messung kognitiver Fähigkeiten abzuleiten, da die faktorielle Struktur auch immer von der verwendeten Aufgaben- und Personenstichprobe abhängig ist (Vernon, 1964). Ein Faktormodell, das mit dem konzeptuellen Modell von Vernon vereinbar ist, ist in Abbildung 7b dargestellt. Daraus ist ersichtlich, dass interindividuelle Unterschiede bei Maßen mathematischer Fähigkeiten durch stoffgebietspezifische Fähigkeiten, generelle mathematische Fähigkeit, verbal-numerische Fähigkeit und allgemeine kognitive Fähigkeit erklärt werden.

(a) Konzeptuelle Repräsentation



(b) Eine mögliche Modellspezifikation



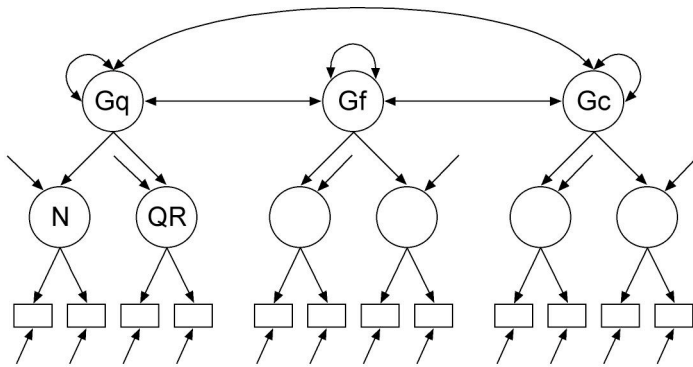
g = allgemeine kognitive Fähigkeit, v:ed = verbal-numerische Fähigkeit, k:m = praktisch-mechanisch-räumlich-physikalische Fähigkeit, n = generelle mathematische Fähigkeit, Geo = Geometrie, Ari = Arithmetik.

Abbildung 7: Vernons hierarchisches Modell der Intelligenz

*Modell fluider und kristalliner Fähigkeiten.* Im Gegensatz zur britischen Forschung basiert die amerikanische Forschung zur Struktur kognitiver Fähigkeiten auf dem oben beschriebenen faktoranalytischen *Bottom-up*-Ansatz. Ein daraus resultierendes, sehr populäres hierarchisches Modell ist die Theorie kristalliner (Gc) und fluider (Gf) Fähigkeiten, die von Cattell (1941) begründet wurde. Fluide Fähigkeit ist nach Cattell die Fähigkeit zum induktiven und deduktiven Schlussfolgern bei *neuartigen* Problemen. Kristalline Fähigkeiten spiegeln sich beim Lösen *vertrauter* Probleme wider und repräsentieren Fähigkeiten und Wissen, die durch den Schulbesuch und die Sozialisation in der jeweiligen Kultur entstehen (siehe auch Abschnitt 5.2.1, Investmenttheorie).

Seit Mitte der 1960er Jahre haben Cattell und Horn und später in erster Linie Horn das zweifaktorielle Gf-Gc-Modell sukzessive auf insgesamt neun (breite) kognitive Fähigkeiten erweitert (*extended Gf-Gc-Theory*, vgl. Abb. 8; Horn & Noll, 1997). Die postulierten kognitiven Fähigkeiten sind, neben Gf und Gc, Verarbeiten visueller Information (Gv), Kurzzeitgedächtnis (Gsm), Langzeitgedächtnis (Glr), Informationsverarbeitungsgeschwindigkeit (Gs), Entscheidungsgeschwindigkeit (CDS), Verarbeiten auditorischer Informationen (Ga) und generelle mathematische Fähigkeit (Gq, die Horn und Noll als *quantitative knowledge* bezeichnen).

Die generelle mathematische Fähigkeit ist von besonderem Interesse für die vorliegende Arbeit: Nach Horn und Noll (1997, S. 69) erfordern Aufgaben zur Messung von Gq Verständnis und Anwendung mathematischer Konzepte und Fertigkeiten. Interindividuelle Unterschiede bei Aufgaben zur Messung mathematischer Fähigkeiten werden – bei einer strengen Interpretation des Gf-Gc-Modells – durch einen Faktor, der generelle mathematische Fähigkeit repräsentiert, und weitere mathematikspezifische Primärfähigkeiten (Abb. 8, N und QR) erklärt. Der Einfluss von Gq wird dabei über die Primärfähigkeiten vermittelt. Horn und Noll nehmen in der aktuellen Fassung des Gf-Gc-Modells *nicht* an, dass eine allgemeine kognitive Fähigkeit existiert. Daher sind anstelle eines generellen Faktors nur latente Interkorrelationen in Abbildung 8 eingezeichnet. Jedoch machen die beiden Autoren keine Aussage darüber, wie diese Interkorrelationen interpretiert werden können.



Gq = generelle mathematische Fähigkeit, Gf = fluide Fähigkeit, Gc = kristalline Fähigkeit, N = Rechenfertigkeit, QR = quantitatives Reasoning. Um die Abbildung übersichtlicher zu gestalten, sind nicht alle von Horn und Noll (1997) identifizierten Fähigkeiten (siehe Text) eingezeichnet.

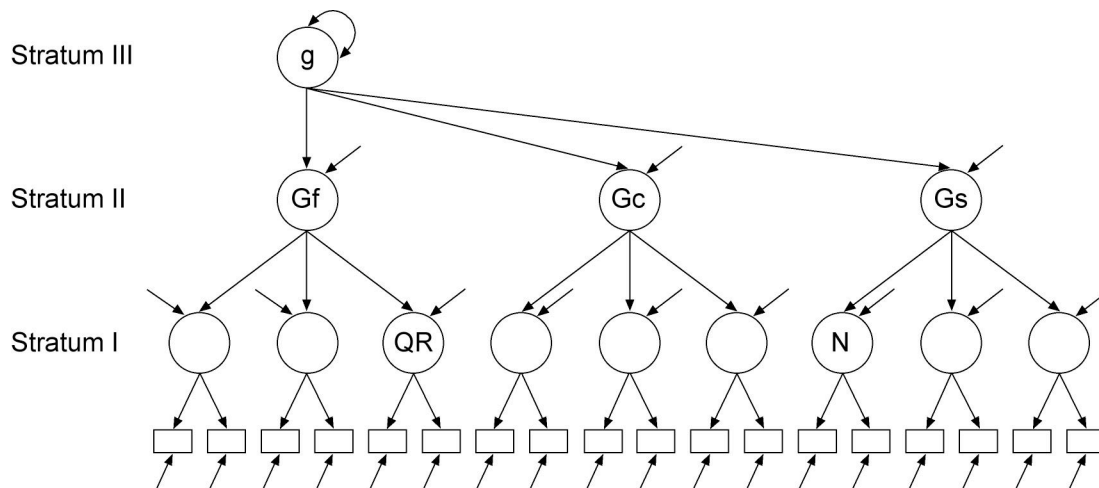
Abbildung 8: Modell fluiden und kristallinen Fähigkeiten (Horn & Noll, 1997)

*Carrolls Drei-Stratum-Theorie.* Im Gegensatz zum Gf-Gc-Modell ist die allgemeine kognitive Fähigkeit ein wesentlicher Bestandteil der Drei-Stratum-Theorie von Carroll (1993), die ebenfalls auf dem faktoranalytischen *Bottom-up*-Ansatz beruht. Dieses theoretische Modell wird von vielen Forschern als die umfassendste Taxonomie kognitiver Fähigkeiten betrachtet (z.B. Gustafsson & Undheim, 1996; Lubinski, 2000; Sternberg, 2003). Dies beruht nicht zuletzt auf der empirischen Datengrundlage, die zu Carrolls Strukturmodell führte: Carroll reanalytierte 461 Datensätze, die in früheren Studien zur Struktur kognitiver Fähigkeiten verwendet wurden. Die Daten stammen dabei von insgesamt über 130.000 Personen mit einem Altersspanne von 6 Monaten bis 71 Jahren, aus 19 verschiedenen Nationen.

Ein zentrales Ergebnis von Carrolls Analyse ist die Unterscheidung von drei *Strata* von Fähigkeitskonstrukten (Abb. 9). Die Strata beziehen sich auf die Generalität (mit den Abstufungen Stratum III: allgemein, Stratum II: breit, Stratum I: eng) der jeweiligen Fähigkeiten. Auf *Stratum III* befindet sich die allgemeine kognitive Fähigkeit (g).

Die allgemeine kognitive Fähigkeit zeichnet sich nach Carroll dadurch aus, dass die darauf ladenden Faktoren und Aufgaben in ihren Anforderungen sehr heterogen sind und komplexe kognitive Prozesse lösungsrelevant sind. Auf *Stratum II* unterscheidet Carroll acht breite Fähigkeiten, die in ihrer Beschreibung mit denen der Gf-Gc-Theorie nahezu identisch sind (Flanagan, McGrew & Ortiz, 2000; McGrew, 1997). Auf *Stratum I* befinden sich mehr als 30 Primärfähigkeiten, wovon zwei eindeutig mathematische Fähigkeitskonstrukte darstellen: quantitatives Reasoning (QR) und Rechenfertigkeit (N). Während für quantitatives Reasoning die problemorientierte Anwendung mathematischer Konzepte notwendig ist, ist nach Carroll Rechenfertigkeit durch die Fähigkeit im Umgang (Addition, Subtraktion, Division, Multiplikation) mit Zahlen bei meist einfach zu lösenden Aufgaben charakterisiert (Tab. 1; vgl. Carroll, 1993). Da nach Carroll bei Aufgaben zur Messung der Rechenfertigkeit die kognitiven Anforderungen an die mentale Informationsverarbeitungsgeschwindigkeit dominieren, ist Rechenfertigkeit der Fähigkeit „Informationsverarbeitungsgeschwindigkeit“ (Carroll nennt diese Fähigkeit *cognitive speed*) hierarchisch untergeordnet.

Die Leistung bei mathematischen Aufgaben wird im Rahmen der Drei-Stratum-Theorie durch interindividuelle Unterschiede bei drei kognitiven Fähigkeiten beeinflusst. Zum Beispiel werden die interindividuellen Unterschiede bei Aufgaben zur Messung von quantitativem Reasoning durch die Fähigkeit zum quantitativen Reasoning, durch die fluide Fähigkeit und durch die allgemeine kognitive Fähigkeit erklärt. Der Einfluss von Fähigkeiten, die höheren Strata zugeordnet sind, wird dabei durch die Fähigkeiten vermittelt, die niedrigeren Strata zugeordnet sind. So wird beispielsweise der Einfluss der allgemeinen kognitiven Fähigkeit auf die Maße zur Messung von quantitativem Reasoning über die fluide Fähigkeit und quantitatives Reasoning vermittelt.



g = allgemeine kognitive Fähigkeit, Gf = fluide Fähigkeit, Gc = kristalline Fähigkeit, Gs = Informationsverarbeitungsgeschwindigkeit, QR = quantitatives Reasoning, N = Rechenfertigkeit. Um die Abbildung übersichtlich zu gestalten, sind nicht alle von Carroll identifizierten Fähigkeiten eingezeichnet.

Abbildung 9: Carrolls (1993) Drei-Stratum-Theorie

*Das Berliner Intelligenzstrukturmodell (BIS-Modell).* Die bisher vorgestellten Modelle basieren in erster Linie auf empirisch induktivem Vorgehen. Dies trifft auch auf die Entwicklung des BIS-Modells zu. Im Vergleich zu den anderen Modellen ist die Entwicklung des BIS-Modells aber auch durch theoriegeleitete deduktive Phasen gekennzeichnet, in denen die Modellannahmen geprüft wurden (Jäger, 1967, 1982, 1984; zusammenfassend Jäger u.a., 1997). In der gegenwärtigen Fassung des BIS-Modells gehen Jäger u.a. (1997) von drei theoretischen Kernannahmen aus:

- (a) Fähigkeitskonstrukte können hinsichtlich des Inhalts und der kognitiven Operationen klassifiziert werden (*Multimodalität*, Abb. 10a). Zu den kognitiven *Operationen* gehören Merkfähigkeit, Einfallsreichtum, Bearbeitungsgeschwindigkeit und Verarbeitungskapazität (das ist vornehmlich die Fähigkeit zum schlussfolgernden Denken). Zu den *Inhalten* gehören figuralbildhaftes Denken, sprachgebundenes Denken und eine generelle mathematische Fähigkeit, die Jäger u.a. als numerisches (zahlengebundenes) Denken bezeichnen. Die operativen und inhaltsgebundenen Fähigkeitskonstrukte interkorrelieren positiv und ermöglichen so die Extraktion eines Faktors, der die allgemeine kognitive Fähigkeit repräsentiert. Letzere nennen Jäger u.a. (1997) „Allgemeine Intelligenz“.

- (b) Die Fähigkeitskonstrukte im BIS-Modell sind hierarchisch organisiert und können so unterschiedlichen Generalitätsebenen zugeordnet werden (*hierarchischer Aufbau*). Die „Allgemeine Intelligenz“ ist das generellste Konstrukt, die operativen und inhaltsgebundenen Fähigkeiten sind spezifischere Konstrukte.
- (c) Bei einer Aufgabe zur Messung kognitiver Fähigkeiten sind alle (im Modell vorgesehenen) intellektuellen Fähigkeiten beteiligt, allerdings mit deutlich unterschiedlichen Gewichten (*multiple Bedingtheit*). In einem korrespondierenden Faktormodell können Aufgaben daher auf einem operativen und inhaltsgebundenen Faktor sowie auf einem Faktor laden, der allgemeine kognitive Fähigkeit repräsentiert (Abb. 10b).

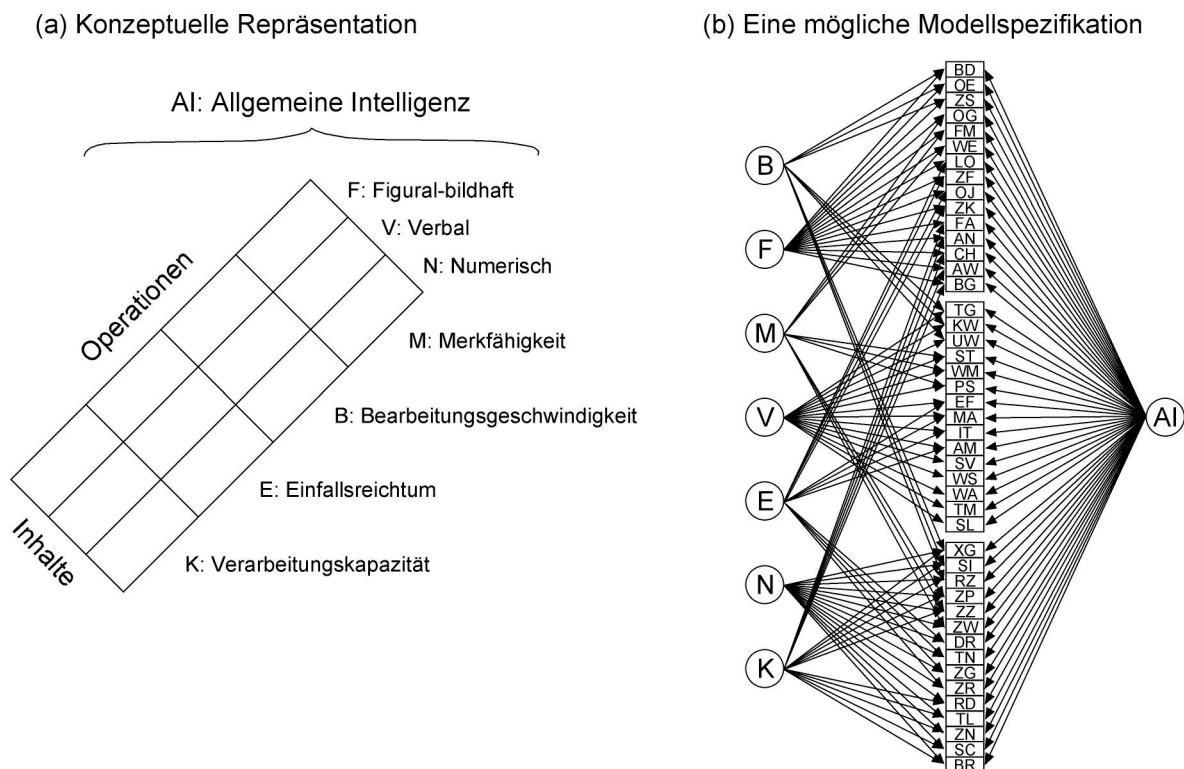


Abbildung 10: Das Berliner Intelligenzstrukturmodell (nach Brunner & Süß, 2005).

Welche Rolle spielen mathematische Fähigkeiten im BIS-Modell? Man könnte meinen, dass im BIS-Modell die „Kreuzung“ aus Operationen und Inhalte (z.B. die Kreuzung von Verarbeitungskapazität und numerischer Fähigkeit) eigene Fähigkeitskonstrukte konstituiert. Somit könnte man numerische Merkfähigkeit, numerische Bearbeitungsgeschwindigkeit (dies entspricht Rechenfertigkeit in den anderen Modellen), numerischen Einfallsreichtum und numerische Verarbeitungskapazität (dies entspricht quantitativem Reasoning in den anderen Modellen) unterscheiden. Darüber hinaus könnte man aufgrund der Annahme des hierarchischen Aufbaus eine generelle mathematische Fähigkeit (diese entspricht konzeptuell weitestgehend G<sub>q</sub> im G<sub>f</sub>-G<sub>c</sub>-Modell und n im Modell von Vernon) konzeptionalisieren. Für die aktuelle Form des BIS-Modells gehen Jäger u.a. (1997) aber nicht davon aus, dass die vier „Fähigkeitskreuzungen“ eigenständige Fähigkeitskonstrukte repräsentieren. Für die Autoren des BIS-Tests stellt die Möglichkeit zur Kreuzklassifikation lediglich einen Orientierungsrahmen dar, in den Testaufgabentypen zur Messung kognitiver Fähigkeiten eingeordnet werden können. Streng interpretiert lässt das BIS-Modell

somit nur eine generelle mathematische Fähigkeit zu. Die Leistung bei mathematischen Aufgaben wird im Rahmen des BIS-Modells also im Wesentlichen durch drei Varianzquellen beeinflusst. Dies sind die jeweilige operative Fähigkeit, die generelle mathematische Fähigkeit und die allgemeine kognitive Fähigkeit.

*Zusammenfassung.* Zusammenfassend kann man festhalten, dass die dargestellten Modelle einige strukturelle Gemeinsamkeiten aufweisen, sich aber auch in wesentlichen Aspekten unterscheiden. Gemeinsam ist allen aktuelleren Modellen (womit die Modelle von Vernon, Gf-Gc, Carroll, BIS-Modell gemeint sind), dass sie einen hierarchischen Aufbau kognitiver Fähigkeiten annehmen und dass mathematische Fähigkeiten berücksichtigt werden. Allerdings unterscheiden sich die Modelle konzeptuell in drei Gesichtspunkten:

- (a) *Die Existenz „allgemeiner kognitiver Fähigkeit“.* Ein Hauptunterschied zwischen den Strukturmodellen liegt darin, ob an der Spitze der Hierarchie eine allgemeine kognitive Fähigkeit (g) postuliert wird (Vernon, Carroll, BIS-Modell) oder nicht (Gf-Gc-Theorie) (siehe auch Flanagan u.a., 2000).
- (b) *Multiple Bedingtheit von Maßen kognitiver Fähigkeiten.* Ein weiterer Unterschied besteht darin, ob angenommen wird, dass die Aufgabenperformanz von mehreren Fähigkeiten abhängig und somit multipel bedingt ist. Dies ist offensichtlich im BIS-Modell der Fall. Weniger offensichtlich ist dies bei den Modellen von Carroll und Vernon. Jedoch können bei der hierarchischen Gruppenfaktorentechnik, die Vernon verwendete, alle Aufgaben gleichzeitig auf mehreren Faktoren laden. Damit setzten sich die Aufgabenvarianzen ebenfalls additiv aus mehreren Faktoren zusammen. Carrolls Ableitung des Modell basiert auf Schmid-Leiman orthogonalisierten Faktoren (Schmid & Leiman, 1957), bei denen die Aufgaben ebenfalls auf mehreren Faktoren laden können (Carroll, 1993, S. 90). In der Gf-Gc-Theorie hingegen wird angenommen, dass die jeweiligen Aufgaben nur auf einem Faktor im Sinne einer Einfachstruktur laden (Horn & Noll, 1997, S. 69) und sich die Aufgabenvarianzen somit nicht aus mehreren Faktoren zusammensetzten (siehe auch Gustafsson & Undheim, 1996).
- (c) *Konzeption mathematischer Fähigkeiten.* Weiterhin wird in allen Modellen explizit oder implizit angenommen, dass mathematische Fähigkeit (bei Carroll jedoch nur die Fähigkeit quantitatives Reasoning) hierarchisch organisiert ist. Unterschiede bestehen jedoch in der Definition des Konstrukts „mathematische Fähigkeit“, die sich darin widerspiegeln, ob mathematische Fähigkeit eher nach inhaltlichen oder operativen Gesichtspunkten konzeptionalisiert wird.
  - Vernon unterscheidet eine generelle numerische Fähigkeit. Diese ist stoffgebietspezifischen mathematischen Fähigkeiten hierarchisch übergeordnet.
  - In der Gf-Gc-Theorie wird angenommen, dass eine breite mathematische Fähigkeit Gq existiert, die sowohl problemorientiertes Anwenden *und* Verständnis mathematischer Konzepte *als auch* mathematische Fertigkeiten umfasst (vgl. Horn & Noll, 1997, S. 66). Dies entspricht einer operativen Konzeptionalisierung mathematischer Fähigkeit.
  - Im BIS-Modell konstituiert sich eine generelle mathematische Fähigkeit aus vier operativen numerischen Facetten.
  - Im Unterschied zu den anderen Strukturmodellen geht Carroll nicht davon aus, dass eine generelle mathematische Fähigkeit existiert (siehe auch Flanagan u.a., 2000). Aus den



Annahmen seiner Drei-Stratum-Theorie kann man aber folgern, dass Carroll eine integrative Sicht einnimmt. Einerseits differenziert er mathematische Fähigkeiten operativ nach Anforderungen an quantitatives Reasoning und Rechenfertigkeit. Andererseits geht Carroll davon aus, dass quantitatives Reasoning die Fähigkeit zur problemorientierten Anwendung von Konzepten der mathematischen Stoffgebiete Arithmetik, Algebra und Geometrie (vgl. Tab. 1; Carroll, 1993, S. 213) widerspiegelt. Letzteres stellt eine inhaltliche Differenzierung dar.

Jedoch muss bei der Unterscheidung mathematischer Fähigkeiten nach inhaltlichen oder operativen Gesichtspunkten eine Einschränkung angemerkt werden. In der bisherigen Forschung zur Ableitung der Strukturmodelle wurden in der Regel Aufgaben aus allen vier gängigen mathematischen Stoffgebieten (Arithmetik, Algebra, Geometrie und Stochastik) *nicht* ausgewogen eingesetzt. Meist dominierten Aufgaben aus der Arithmetik, und entsprechende Fähigkeiten werden daher als „numerische“ (vgl. n im Modell von Vernon oder N im BIS-Modell) oder „quantitative“ Fähigkeiten (vgl. Gq im Gf-Gc-Modell oder QR im Modell von Carroll) und nicht als „mathematische“ Fähigkeiten bezeichnet. Inwiefern die theoretischen Annahmen aus den Strukturmodellen auf andere mathematische Stoffgebiete übertragbar sind, ist damit – empirisch gesehen – nicht so gut abgesichert wie für Aufgaben aus der Arithmetik. Andererseits erscheint es aber gerechtfertigt, von der Gültigkeit der Strukturmodelle zumindest für die gängigen mathematischen Stoffgebiete auszugehen. Strukturmodelle streben das Ziel an, das *celestial white* (Horn, 1988) kognitiver Fähigkeiten taxonomisch zu ordnen (Gustafsson & Undheim, 1996). Daher sollten die Annahmen, die in den Strukturmodellen gemacht werden, auch für Aufgaben aus den mathematischen Stoffgebieten Algebra, Geometrie und Stochastik zutreffen. Nicht zuletzt wird dieser Anspruch auch von Horn und Noll explizit so formuliert: „Quantitative knowledge (Gq), measured in tasks requiring understanding and application of the concepts and skills of *mathematics*.“ [Hervorhebung M.B.] (Horn & Noll, 1997, S. 69)

Tabelle 2 fasst nochmals die diskutierten Unterschiede zwischen den Strukturtheorien zusammen. Die diskutierten Unterschiede zwischen den Strukturmodellen kognitiver Fähigkeiten haben direkte Implikationen, wie interindividuelle Unterschiede bei Maßen mathematischer Schülerleistung erklärt werden können. Die zentralen Hypothesen (für die konkrete Formulierung siehe Kap. 7), die im Rahmen der ersten Forschungsfrage behandelt werden, beschäftigen sich daher mit zwei Fragen:

- Erklären primär stoffgebietspezifische Fähigkeiten oder operative mathematische Fähigkeiten die interindividuellen Unterschiede bei Aufgaben zur Messung mathematischer Schülerleistung?
- Welche Rolle spielen generelle mathematische Fähigkeiten und die allgemeine kognitive Fähigkeit, bei der Erklärung interindividueller Unterschiede bei Maßen mathematischer Schülerleistung?

Tabelle 2: Zentrale Unterschiede zwischen den Strukturtheorien kognitiver Fähigkeiten

Strukturtheorie	Existenz „allgemeiner kognitiver Fähigkeit“?	Multiple Bedingtheit von Maßen kognitiver Fähigkeiten?	Konzeption mathematischer Fähigkeiten
Vernon	Ja	Ja	Stoffgebiete
Gf-Gc-Modell	Nein	Nein	Operative Fähigkeiten
Drei-Stratum-Theorie	Ja	Ja	Stoffgebiete und operative Fähigkeiten
BIS-Modell	Ja	Ja	Operative Fähigkeiten

Gf = fluide Fähigkeiten, Gc = kristalline Fähigkeiten, BIS-Modell = Berliner Intelligenzstrukturmodell.

Antworten auf die beiden Fragen betreffen das erste Hauptziel dieser Arbeit: eine Anbindung mathematischer Schülerleistung an differenzialpsychologische kognitive Fähigkeitskonstrukte.

Natürlich stellen die vorgestellten Modelle nur einen Ausschnitt aus den in der Forschungsliteratur diskutierten Modellen dar (Brody, 2000; Gustafsson & Undheim, 1996; Li & Schmiedek, 2001; Wilhelm, 2000). Strukturmodelle, auf die nicht eingegangen wurde, sind:

- Das Modell von Royce (1973), da darin mathematische Fähigkeiten nicht berücksichtigt werden.
- Die Annahme von über 120 verschiedenen Facetten im *Structure of Intellect Modell* von Guilford (1967) und die verwendete faktoranalytische Methode (Horn & Knapp, 1973) lassen an der Modellgültigkeit zweifeln (zusammenfassend Carroll, 1993). Ferner wird die Unterscheidung operativer und inhaltlicher Facetten bereits im Kontext des BIS-Modells diskutiert.
- Die inhaltsgebundene Facettenstruktur des Modells von Snow und Kollegen (Marshalek, Lohman & Snow, 1983; Snow, Kyllonen & Marshalek, 1984; Snow & Lohman, 1989) wie auch die Tatsache, dass sich Aufgaben zur Messung kognitiver Fähigkeiten hinsichtlich der Ladungen auf einem Generalfaktor unterscheiden können, sind im Rahmen des BIS-Modells angesprochen worden.
- Gleiches gilt auch für das Modell von Guttman (1965; Guttman & Levy, 1991), in dem ähnliche Annahmen wie im Modell von Snow und Kollegen getroffen werden.

Bei der Darstellung der Strukturmodelle in diesem Abschnitt wurde ausgespart, wie die Struktur der kognitiven Fähigkeiten entstehen kann. Dies ist Gegenstand des nächsten Abschnitts.

## 4.3 Integration und Differenzierung kognitiver Fähigkeiten

### 4.3.1 Allgemeine „Mechanismen“

Warum können interindividuelle Unterschiede bei Aufgaben zur Messung kognitiver Fähigkeiten durch Faktoren erklärt werden? Was führt dazu, dass Aufgaben stärker oder schwächer miteinander korrelieren? Eine Zunahme der Korrelationen ist indikativ für eine Integration der kognitiven

Fähigkeiten: Die Korrelation zwischen Faktoren nimmt zu. Die Abnahme der Korrelationen zwischen Aufgaben zur Messung kognitiver Fähigkeiten ist indikativ für eine Differenzierung kognitiver Fähigkeiten: Die Korrelation zwischen Faktoren nimmt ab und die jeweiligen Faktorvarianzen nehmen (unabhängig voneinander) zu. Letzteres bedeutet auch, dass die *Heterogenität* der kognitiven Fähigkeiten in einer bestimmten Personenpopulation zunimmt. Die Überlegungen und Befunde zur Differenzierung kognitiver Fähigkeiten sind daher sehr relevant vor dem Hintergrund der zweiten Forschungsfrage, in der es unter anderem darum geht, die Heterogenität der kognitiven Fähigkeiten mathematischer Schülerleistung in Abhängigkeit der Schulform zu analysieren.

Die Fragen nach Integration und Differenzierung kognitiver Fähigkeiten sind weder neu (z.B. Tryon, 1935), noch gibt es einen universellen Mechanismus, wie die zahlreichen in der Literatur diskutierten Ansätze belegen (Anastasi, 1970, 1986; Carroll, 1966; Ferguson, 1954, 1956; Vernon, 1965). Nachfolgend werden drei allgemeine Ansätze dargestellt, die zur Erklärung herangezogen werden: kognitive Ressourcen und Informationsverarbeitungsprozesse, Transfereffekte und Lernumwelten. Diese Ansätze sind nicht spezifisch für mathematische Fähigkeiten oder gar mathematische Schülerleistung formuliert worden. Jedoch lassen sich die darin formulierten Annahmen problemlos darauf übertragen.

*Kognitive Ressourcen und Informationsverarbeitungsprozesse.* Die zentrale Annahme dieser Ansätze ist, dass bestimmte kognitive Ressourcen oder Informationsverarbeitungsprozesse zentral für die Performanz bei einer Klasse von Aufgaben sind. Interindividuelle Unterschiede in den kognitiven Ressourcen oder den Informationsverarbeitungsprozessen führen dazu, dass die Performanz bei diesen Aufgaben interkorreliert. Diese Annahme findet sich bereits bei Spearman (z.B. 1927), der davon ausgeht, dass die mentalen Prozesse *eduction of relations*, *eduction of correlates* und *apprehension of experience* dem Faktor *g* zu Grunde liegen.<sup>10</sup>

Während die von Spearman postulierten Prozesse auf die allgemeine kognitive Fähigkeit (in Form eines Generalfaktors) fokussieren, ist Tryons (1935) Stichprobenansatz genereller und kann auch das Zustandekommen von verschiedenen kognitiven Fähigkeiten (in Form von Gruppenfaktoren wie z.B. bei Thurstone) erklären. Tryon nimmt an, dass die Interkorrelation von Aufgaben maßgeblich dadurch bestimmt ist, inwieweit die gleichen „psychologischen Komponenten“ zur Bearbeitung zweier Aufgaben lösungsrelevant sind. Die Vorgabe von Testaufgaben vergleicht Tryon mit einer Stichprobenziehung aus dem „Universum lösungsrelevanter psychologischer Komponenten“. Je größer die Schnittmenge der Stichprobenräume zweier Aufgaben ist, aus denen die psychologischen Komponenten „gezogen“ werden, desto höher ist die zu erwartende Aufgabeninterkorrelation.

Diese Überlegung greifen zum Beispiel Snow und Kollegen (Snow u.a., 1984) auf, um das Zustandekommen einer Radexstruktur kognitiver Fähigkeiten zu erklären. Eine Radexstruktur impliziert hierbei, dass sich Aufgaben zur Messung kognitiver Fähigkeiten nach der inhaltlichen Anforderung (indiziert durch die Ladung auf den inhaltlichen Faktoren) und nach den Anforderungen an die Komplexität der Informationsverarbeitungsprozesse (indiziert durch die Ladung

---

<sup>10</sup> *Eduction of relations* ist die Wahrnehmung einer Beziehung zwischen zwei oder mehreren Wissens-elementen (z.B. dass „gut“ und „böse“ einen Gegensatz bilden, Wilhelm, 2000, S. 34). *Eduction of correlates* bedeutet, dass aus der Kenntnis eines Wissens-elementes und der Beziehung zu einem zweiten Wissens-element das zweite Wissens-element erschlossen werden kann (z.B. von „gut“ und der Relation „Gegensatz“ auf „böse“ zu schlussfolgern, Wilhelm, 2000). Der dritte Prozess *apprehension of experience* kann aus heutiger Sicht wohl unter dem breiten Begriff „Metakognition“ subsumiert werden (Carroll, 1993, S. 637).

auf dem Faktor für allgemeine kognitive Fähigkeit) unterscheiden. Snow u.a. (1984) betrachten dabei das kognitive System einer Person als semantisches Netzwerk mit einer Vielzahl von „Komponenten“ (Informationsverarbeitungsprozesse, Schemata, Pläne oder Wissensstrukturen). Die erfolgreiche Bearbeitung einer Aufgabe erfordert dabei die Verfügbarkeit einer Vielzahl dieser Komponenten, die zudem richtig sequenziert werden müssen. Personen unterscheiden sich darin, wie gut die von ihnen produzierten Komponentensequenzen den Aufgabenanforderungen entsprechen. Wenn die Komponentensequenzen und Aufgabenanforderungen hinreichend gut übereinstimmen, wird eine Aufgabe erfolgreich gelöst. Hohe Interkorrelationen zwischen zwei Aufgaben resultieren dann, wenn ähnliche Stichproben von Komponentensequenzen für beide Aufgaben lösungsrelevant sind.

Als eine aktuelle Weiterführung der Überlegungen von Snow u.a. kann das 4-Quellen-Modell betrachtet werden (z.B. Kyllonen, 1994; Kyllonen & Christal, 1990): Zentral ist im 4-Quellen-Modell, dass das Zustandekommen interindividueller Unterschiede bei kognitiven Aufgaben durch interindividuelle Unterschiede bei vier zentralen Komponenten (Quellen) der kognitiven Architektur erklärt wird: Interindividuelle Unterschiede können dabei in Tiefe und Breite des deklarativen Wissens, in der Tiefe und Breite des prozeduralen Wissens, in der Kapazität des Arbeitsgedächtnisses (als Fähigkeit zum simultanen Speichern und Verarbeiten von Informationen) und in der Geschwindigkeit von Informationsverarbeitungsprozessen bestehen (Kyllonen & Christal, 1990). Deklaratives und prozedurales Wissen kann dabei noch weiter nach der jeweiligen Wissensdomäne (verbal, quantitativ, figural-räumlich, z.B. Kyllonen, 1994) binnendifferenziert werden. Aufgabeninterkorrelationen resultieren, weil für eine Klasse von Aufgaben dieselben kognitiven Komponenten lösungsrelevant sind.

Man kann die Annahme, dass kognitive Ressourcen oder Informationsverarbeitungsprozesse zur Interkorrelation der Aufgaben führen, auch aus einer lernpsychologischen Perspektive betrachten: Kognitive Ressourcen und Informationsverarbeitungsprozesse werden von einer Aufgabe auf die nächste transferiert.

*Transfereffekte.* Ein Vertreter dieses Ansatzes ist Ferguson (1954, 1956). Ausgangspunkt von Ferguson ist einerseits die Beobachtung aus der Lernpsychologie, dass Lernkurven sich asymptotisch einer Performanzgrenze nähern, und andererseits die Annahme zeitlicher Stabilität psychometrisch definierter Fähigkeiten. Ferguson versucht diese beiden Aspekte zu integrieren, indem er kognitive Fähigkeiten als überlerntes Verhalten betrachtet: Die erzielte Leistung bei kognitiven Aufgaben resultiert für Ferguson durch zeitlich lang andauerndes „Training“ bei gleichen oder ähnlichen kognitiven Aufgaben. Die Aufgabenleistung entspricht damit der asymptotischen Performanz am Ende der Trainingsphase. Diese asymptotische Performanz ist zeitlich weitestgehend stabil, was die zeitliche Stabilität von Fähigkeiten erklärt.

Darüber hinaus geht Ferguson davon aus, dass die Fähigkeit, eine bestimmte Klasse von Aufgaben zu lösen, auf ähnliche Aufgaben transferiert und somit die Performanz positiv beeinflusst (siehe auch Carroll, 1966; Cattell, 1987). Dies führt zur Interkorrelation zwischen Aufgaben. Je weiter die Transfereffekte sind, desto „breiter“ ist die korrespondierende kognitive Fähigkeit, die diese Interkorrelationen erklärt. Was tatsächlich gelernt wird, hängt natürlich in hohem Maße von den bearbeiteten Aufgaben ab.

*Lernumwelten.* Die Umwelt einer Person (z.B. das familiäre Umfeld oder die Schule) hat damit entscheidenden Einfluss auf die Struktur der kognitiven Fähigkeiten in einer Personenpopulation. In der Lernumwelt werden Aufgaben vorgegeben, die bearbeitet werden können. Somit bestimmt

die Umwelt vermittelt über die eingesetzten Aufgaben maßgeblich die Lernerfahrungen einer Person (siehe auch Anastasi, 1970, 1986; Carroll, 1966; Tryon, 1935). Zentral ist hierbei der Gedanke, dass sich die Möglichkeit, bestimmte Lernerfahrungen zu machen, zwischen Umwelten unterscheidet. Carroll (1993) diskutiert diese differenziellen Lernerfahrungen im Rahmen eines Gedankenexperiments: Schüler werden zufällig auf vier Gruppen aufgeteilt. Jede Gruppe wird in einem anderen Fachgebiet unterrichtet, das den Schülern vor der Untersuchung unbekannt war. Carroll geht davon aus, dass nach Ende der Unterrichtsperiode bei Vorgabe von zwei oder drei Tests pro Fachgebiet Faktoren für jedes Fachgebiet extrahiert werden können. Diese Faktoren repräsentieren die kognitiven Fähigkeiten zur Bearbeitung der jeweiligen Fachgebiete. Carroll impliziert in seinem Gedankenexperiment, dass nur die Schüler in einer bestimmten Gruppe die Möglichkeit haben, die Inhalte des jeweiligen Fachgebiets zu lernen. Dies führt (bei Analyse der Daten aller vier Gruppen) zu interindividuellen Unterschieden bei der Performanz korrespondierender Testaufgaben, die das Beherrschen der Lerninhalte des jeweiligen Fachgebiets erfassen. Wenn alle Testaufgaben auf die Lerninhalte dieses Fachgebiets abzielen, korreliert die Performanz bei den Testaufgaben und ermöglicht so die Extraktion eines Faktors. Wenn dies für alle vier Schülergruppen zutrifft, können bei Analyse der gesamten Stichprobe vier Faktoren für jedes Fachgebiet extrahiert werden.

Für Carroll führt jede Kultur oder Gesellschaft solche „Experimente“ durch, da Menschen unterschiedliche Lernerfahrungen machen. Einerseits lernen Menschen bei freiwillig gewählten Aktivitäten in der Freizeit oder am Arbeitsplatz. Andererseits können diese Lernerfahrungen wie beispielsweise der Schulbesuch gesellschaftlich „verordnet“ sein. Daher ist klar, dass Unterschiede zwischen Lehrplänen oder Unterschiede in den instruktionalen Maßnahmen Einfluss auf die vorgefundene Struktur kognitiver Fähigkeiten haben können. Die Höhe der Interkorrelation und die Breite der Faktoren, die in der Schule erworbene kristalline Fähigkeiten repräsentieren, sind davon abhängig, welche Inhalte in den Lehrplänen vorgesehen sind und wie diese Inhalte im Unterricht implementiert werden (vgl. Carroll, 1966; Cattell, 1987; Vernon, 1964, 1969).

Im Zusammenhang mit der Wirkung von Lernumwelten können mit Baltes, Nesselroade und Cornelius (1978) zwei Prozesse unterschieden werden (Abb. 11), die zur Differenzierung oder Integration (von Faktoren) führen. Das von Carroll beschriebene Gedankenexperiment bezieht sich auf die Differenzierung und beschreibt das Entstehen von Gruppenfaktoren. Die Integration beschreibt hingegen das Zustandekommen eines generellen Faktors.

Im Rahmen einer Simulationsstudie konnten Baltes u.a. (1978) zeigen, dass sowohl Integration als auch Differenzierung aus Umweltfaktoren resultieren können. Entscheidend ist hierbei die Annahme, dass Personen sich darin unterscheiden, welche Umweltfaktoren auf sie wirken. Integration und Differenzierung basieren dabei auf zwei unterschiedlichen Mechanismen:

- (a) Wenn Umweltfaktoren positiven Einfluss auf die Performanz bei allen kognitiven Aufgaben haben, dann resultiert eine allgemeine kognitive Fähigkeit (in Form eines generellen Faktors) und die Interkorrelation zwischen kognitiven Fähigkeiten (repräsentiert durch Gruppenfaktoren) nimmt zu. Baltes u.a. (1978) sprechen hierbei von allgemeinen Umweltunterschieden (*general environmental differences*), die zur Integration führen. In der Simulationsstudie wurden deshalb Kovarianzen bei allen Aufgaben hinzugefügt (in Abb. 11 wird dies durch einen Faktor repräsentiert, der auf alle manifesten Variablen Einfluss hat).

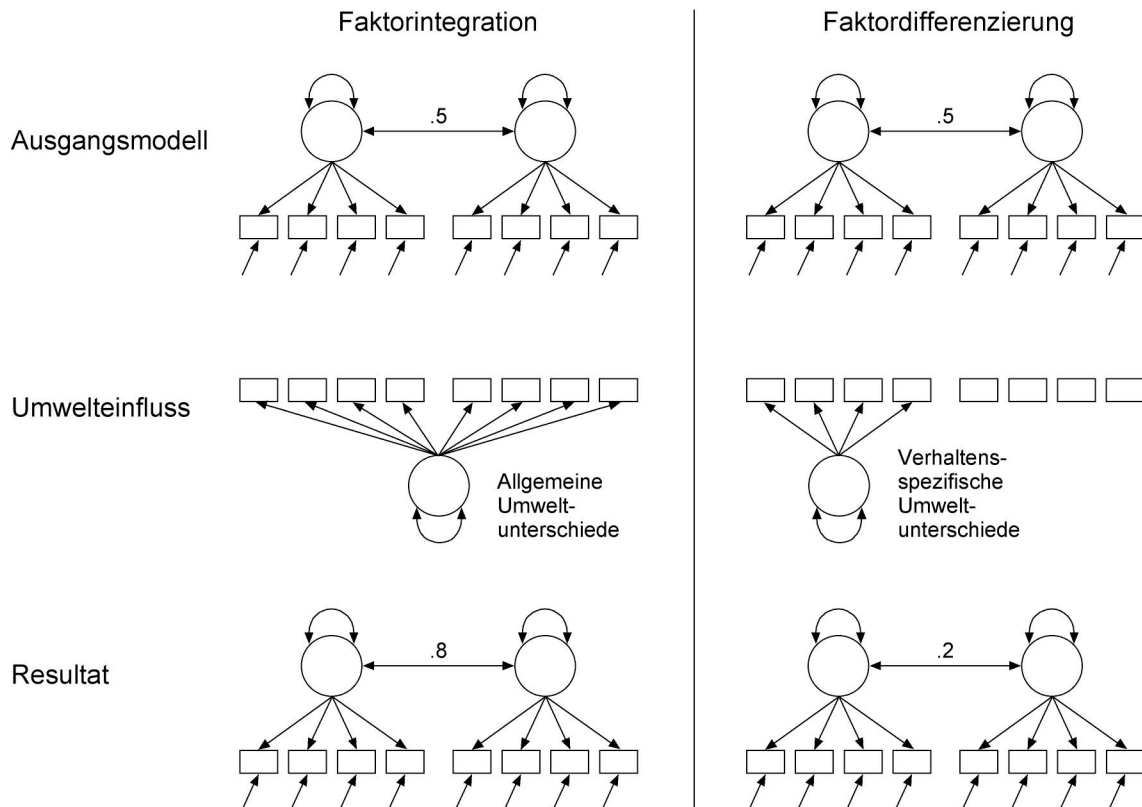


Abbildung 11: Schematische Darstellung von Faktorintegration und Faktordifferenzierung (in Anlehnung an Baltes u.a., 1978)

- (b) Wenn die Umweltfaktoren jedoch nur die Performanz bei den Aufgaben zur Messung einer bestimmten kognitiven Fähigkeit positiv beeinflussen (vgl. das Gedankenexperiment von Carroll, 1993), resultiert eine differenziertere Struktur der kognitiven Fähigkeiten und die Interkorrelationen zwischen den kognitiven Fähigkeiten verringern sich. Baltes u.a. (1978) nennen dies verhaltensspezifische Umweltunterschiede (*behavior-specific environmental differences*), die zur Differenzierung der kognitiven Fähigkeiten führen. Zur Simulation dieses Effekts wurden Kovarianzen nur bei den Aufgaben addiert, die zur Messung eines bestimmten Gruppenfaktors verwendet werden (in Abb. 11 wird dies durch einen Faktor repräsentiert, der nur auf einen Teil der manifesten Variablen Einfluss hat).

*Zusammenfassung.* Mit Blick auf mathematische Schülerleistung können (faktoriell repräsentierte) kognitive Fähigkeiten, die interindividuelle Unterschiede bei Maßen mathematischer Schülerleistung erklären, (mindestens) durch drei „Wirkmechanismen“ zu Stande kommen: Informationsverarbeitungsprozesse, die allen Maßen mathematischer Schülerleistung gemein sind, Transfereffekte mathematischen Wissens oder Unterschiede in der „Lernumwelt“ Schule. Die schulische Lernumwelt wird daher in Abschnitt 5.3 und 5.4 noch weiter beleuchtet, um Hypothesen zu „Schulformunterschieden“ für die zweite Forschungsfrage abzuleiten.

Drei verschiedene Erklärungsansätze zur Entstehung von Faktorstrukturen suggerieren möglicherweise, dass diese weitestgehend unabhängig sind. Jedoch sollte man bedenken, dass dies lediglich drei Perspektiven auf ein und dasselbe Phänomen sind: Wenn Informationsverarbeitungsprozesse, Wissensstrukturen oder kognitive Ressourcen lösungsrelevant für eine Klasse von Aufgaben sind, kann man dies als Transfer dieser kognitiven Komponenten beschreiben. Gleich-

zeitig werden diese kognitiven Komponenten auch in einer bestimmten Lernumwelt erworben oder trainiert.

Die in diesem Abschnitt vorgestellten Erklärungsansätze zur Entstehung von Faktorstrukturen gelten universell für alle Personen. Der nächste Abschnitt beschäftigt sich damit, ob und warum sich Faktorstrukturen für Personen mit hohem und niedrigem Niveau allgemeiner kognitiver Fähigkeit unterscheiden. Damit ist ein Teilaspekt der zweiten Forschungsfrage verbunden: Nimmt die Heterogenität mathematischer Fähigkeiten für Schüler mit hohem und niedrigem Niveau allgemeiner kognitiver Fähigkeit zu? Ist diese Differenzierung abhängig von der besuchten Schulform?

#### 4.3.2 Differenzierungshypothese

Mit Reinert (1970) und Carlstedt (2001) kann man die Studien, die die Interkorrelationen kognitiver Fähigkeiten für Personen unterschiedlicher Fähigkeitsniveaus analysieren, in zwei große Gruppen einteilen: In der ersten Gruppe von Studien werden Personen mit unterschiedlichem *chronologischem Alter* untersucht. In diesen Studien wird explizit oder implizit davon ausgegangen, dass sich das Niveau allgemeiner kognitiver Fähigkeit (bzw. fluider Fähigkeit) zunächst erhöht und dann bei alten und sehr alten Menschen wieder sinkt. Die zweite Gruppe von Studien beschäftigt sich mit Personen *unterschiedlichen Fähigkeitsniveaus* (mit nahezu gleichem chronologischem Alter).

*Chronologisches Alter.* Eine zentrale zu Grunde liegende Hypothese ist in diesen Studien, dass mit zunehmendem chronologischem Alter zunächst eine Differenzierung kognitiver Fähigkeiten beobachtet werden kann. Reinert und Kollegen (Reinert, 1970; Reinert, Baltes & Schmidt, 1965) nennen dies die Altersdifferenzierungshypothese, die für die Altersstufen von 6 bis 20 Jahren gültig sein soll und auf Garrett (1946) zurückgeht. Bei alten und sehr alten Menschen wird hingegen davon ausgegangen, dass mit abnehmendem Niveau allgemeiner kognitiver Fähigkeit die Struktur kognitiver Fähigkeiten dedifferenziert (Reinert, 1970), das heißt, man kann eine Integration der kognitiven Fähigkeiten beobachten. Die Ergebnisse dieser Gruppe von Studien zur Differenzierungs–Dedifferenzierungshypothese waren im Jahr 1965 uneindeutig (Reinert u.a., 1965) und werden heutzutage entweder als *puzzling* (Juan-Espinosa u.a., 2002, S. 396) oder aufgrund von methodischen Problemen (Nesselroade & Thompson, 1995) als „inconclusive, but generally (...) supportive“ (Baltes, Lindenberger & Staudinger, 1998, S. 1080) betrachtet. Da in dieser Arbeit keine entwicklungspsychologische Fragestellung im Vordergrund steht, wird die Literatur hierzu nicht weiter aufgearbeitet (siehe z.B. Bickley, Keith & Wolfle, 1995; Carroll, 1993; Juan-Espinosa u.a., 2002; Li u.a., 2004; Schaie, 1994; Schaie, Maitland, Willis & Intrieri, 1998; Schaie, Willis, Jay & Chipuer, 1989).

*Unterschiedliche Fähigkeitsniveaus.* Bei der zweiten Gruppe von Studien werden die Interkorrelationen kognitiver Fähigkeiten in Abhängigkeit des Fähigkeitsniveaus untersucht. Das chronologische Alter spielt dabei als Untersuchungsvariable eine untergeordnete Rolle oder wird (z.B. durch die Stichprobenbildung) bei den Analysen weitestgehend konstant gehalten oder statistisch kontrolliert. Eine der ersten Untersuchungen hierzu stammt von Spearman (1927): Er verglich die Interkorrelationen von Tests zur Messung kognitiver Fähigkeiten von normal entwickelten Schülern und Schülern mit Lernschwierigkeiten. Ergebnis war, dass die Testinterkorrelationen von Schülern mit Lernschwierigkeiten deutlich höher waren als bei den normal entwickelten Schülern (für eine Reanalyse siehe Deary & Pagliari, 1991). Er konnte also eine Faktordifferenzierung mit

zunehmendem Fähigkeitsniveau wie auch die Dedifferenzierung bzw. Faktorintegration mit abnehmendem Fähigkeitsniveau beobachten.

Die Beobachtung von Spearman geriet jedoch in Vergessenheit und wurde erst in jüngerer Zeit von Detterman und Daniel (1989) „wiederentdeckt“ (Deary & Pagliari, 1991) und mehrmals repliziert (siehe aber auch Fogarty & Stankov, 1995; Hartmann & Reuter, 2006):

- (a) Detterman und Daniel (1989) konnten beim Vergleich von Personen mit unterschiedlichen Fähigkeitsniveaus dreierlei feststellen. Erstens, Tests zur Messung kognitiver Informationsverarbeitungsprozesse interkorrelierten im Durchschnitt höher bei Personen mit niedrigem Fähigkeitsniveau. Zweitens, die Korrelation dieser Tests mit einem Indikator für die allgemeine kognitive Fähigkeit gemessen durch den Wechsler-Intelligenztest für Erwachsene (WAIS-R, z.B. Wechsler, 1981), wie auch drittens die Korrelation von Tests zur Messung kognitiver Fähigkeiten des WAIS-R oder dem Wechsler-Intelligenztest für Kinder (WISC-R, z.B. Wechsler, 1974) waren höher für Personen mit niedrigem Fähigkeitsniveau.
- (b) Legree, Pifer und Grafton (1996) replizierten den Befund von Detterman und Daniel anhand einer Standardisierungsstichprobe des *Armed Services Vocational Aptitude Battery* (ASVAB, Department of Defense, 1984). Eine Besonderheit dieses Ergebnisses ist nach Legree u.a. der Unterschied zwischen den Tests: Der Wechsler-Intelligenztest wurde zur Messung kognitiver Fähigkeiten konstruiert. Die ASVAB stellt nach Legree u.a. (1996, S. 46–47) eine Chimäre zwischen einem Test zur Messung kognitiver Fähigkeiten und einem Schülerleistungstest dar. Daher konnte nach Ansicht der Autoren mit einem Verweis auf eine persönliche Kommunikation mit Detterman (vgl. Legree u.a., 1996, S. 46) nicht von vornherein erwartet werden, dass eine Faktordifferenzierung beobachtet werden kann.
- (c) In einer viel beachteten Arbeit replizierten Deary und Kollegen (Deary u.a., 1996) die Befunde zur Faktordifferenzierung anhand von Daten der *American Differenzial Aptitude Tests* (DAT, Educational Research Centre, 1986) und zwei verschiedenen Altersgruppen unter Verwendung einer ausgefeilten Methode zur zufälligen Ziehung alters- und fähigkeitsstratifizierter Analysestichproben.
- (d) Carlstedt (2001) analysierte mit einer eigens für seine Studie entwickelten Methodik Daten von Armeeanwärtern, die eine Testbatterie zur Messung kognitiver Fähigkeiten (*Swedish Enlistment Battery*) bearbeiteten. Bemerkenswert an den Ergebnissen dieser Studie war, dass die Differenzierung der spezifischen verbalen Fähigkeit und der spezifischen figural-räumlichen Fähigkeit nur für Fähigkeitsgruppen im unteren Spektrum allgemeiner kognitiver Fähigkeiten beobachtet werden konnte, wohingegen im oberen Spektrum die Differenzierung (möglicherweise aufgrund von psychometrischen Mängeln der analysierten Tests) ausblieb.
- (e) Abad und Kollegen (Abad, Colom, Juan-Espinosa & Garcia, 2003) bildeten ihre Analysestichprobe mit derselben Methode wie Deary und Kollegen und konnten deren Befunde für Personen mit einem weiteren Range allgemeiner kognitiver Fähigkeiten (indiziert durch sehr heterogene Bildungsabschlüsse) sowie zwei Altersgruppen replizieren. Datengrundlage bildeten die Normierungsstichprobe des Wechsler-Intelligenztests für Erwachsene (WAIS-III) sowie eine Stichprobe von Studienplatzbewerbern, die eine Testbatterie zur Messung kognitiver Fähigkeiten bearbeitete.



- (f) Jensen (2003) teilte die Normierungsstichproben des WISC-R (Wechsler, 1974) in zwei altershomogenere Gruppen und die Normierungsstichprobe des WAIS-R (Wechsler, 1981) in drei altershomogenere Gruppen ein. Innerhalb der Altersgruppen unterteilte er die Personen nochmals in zwei Gruppen mit hohem und niedrigem Fähigkeitsniveau. Über alle Altersgruppen hinweg konnte er die Faktorintegration für Personen mit niedrigem Fähigkeitsniveau feststellen. Ein weiteres wichtiges Ergebnis war, dass nach Jensen die Subtests der beiden Wechsler-Tests mit den geringsten Ladungen auf g am stärksten von der Faktorintegration betroffen waren. Diese Subtests hatten geringe Ladungen auf g für Personen mit hohem Fähigkeitsniveau und hohe Ladungen auf g für Personen mit geringem Fähigkeitsniveau.

Unabhängig von den differenziellen Befunden zu Subtests ist insgesamt gesehen die Faktorintegration und -differenzierung für Personen mit unterschiedlichem Fähigkeitsniveau ein altersgruppenübergreifender, mehrfach replizierter empirischer Befund, der sehr unterschiedlich erklärt wird (für Zusammenfassungen siehe z.B. Deary u.a., 1996; Jensen, 1998). Nachfolgend werden drei Erklärungsansätze vorgestellt. Spearman's Ansatz wird aufgeführt, da er der Erste war, der versuchte, das Phänomen zu erklären. Die nächsten beiden Ansätze sind aus theoretischer Sicht relevant, um die zweite Forschungsfrage nach Schulformunterschieden der Differenzierung mathematischer Fähigkeiten zu beantworten.

*Spearman's „law of diminishing returns“:* Der historisch gesehen erste Erklärungsansatz stammt von Spearman (1927, S. 217–221), der vom *law of diminishing returns* spricht. Da nach Spearman der Faktor g die Interkorrelationen von Tests zur Messung kognitiver Fähigkeiten erklärt, folgert er daraus, dass mit zunehmendem Niveau von g, der Faktor g weniger wichtig für das Zustandekommen der beobachteten Testleistungen ist. Gleichzeitig geht er davon aus, dass die aufgabenspezifischen Faktoren s mit zunehmendem Niveau von g wichtiger werden. Spearman begründet diese Beobachtung durch Analogien zur Physik und Ökonomie: Da Spearman g als mentale Energie betrachtet, führt nach den Gesetzen zur Energieerhaltung eine Verdopplung der eingesetzten Energie nicht zu einer Verdopplung der Geschwindigkeit eines Objekts. Genauso wenig führt entsprechend des ökonomischen *law of diminishing returns* nach Spearman eine Verdopplung des investierten Kapitals zu einer Verdopplung des Gewinns.

*Investmenttheorie und Fertigkeitserwerb.* Eine im Vergleich zu Spearman psychologisch elaboriertere Erklärung, die nicht auf Analogien zur Physik oder Ökonomie zurückgreift, stammt von Jensen (1998, S. 584). Jensen integriert zur Erklärung von Differenzierung und Integration kognitiver Fähigkeiten die Investmenttheorie von Cattell (1987) und Befunde des Fertigkeitserwerbs (siehe Abschnitt 5.2.1 zu einer ausführlicheren Darstellung von Investmenttheorie und Fertigkeitserwerb): Im Zuge der mentalen Entwicklung wird die allgemeine kognitive Fähigkeit zunehmend in spezialisiertere Fähigkeiten investiert. Mit zunehmender Übung werden diese spezialisierteren Fähigkeiten mehr und mehr automatisiert. Die automatisierbaren Aspekte dieser Fähigkeiten sind dann von der allgemeinen kognitiven Fähigkeit unabhängig. Die Kovariation von Aufgaben zur Messung der automatisierten Aspekte ermöglicht die Extraktion von Gruppenfaktoren, die die spezialisierten kognitiven Fähigkeiten repräsentieren.

Zudem nimmt Jensen an, dass Personen mit höherem Fähigkeitsniveau (im Vergleich zu Personen mit niedrigem Fähigkeitsniveau) mehr Möglichkeiten haben, ihre mentalen Ressourcen in den Erwerb verschiedenartiger Fertigkeiten zu investieren. Dies trägt zur Differenzierung der kognitiven Fähigkeiten bei.

*Schulbildung.* Mit Blick auf die unterschiedlichen Investitionsmöglichkeiten sind die Spekulationen von Abad und Kollegen (2003) zur Interpretation ihrer Ergebnisse interessant: Sind Unterschiede in der Schulbildung für Personen mit unterschiedlichem Fähigkeitsniveau für die Differenzierung der kognitiven Fähigkeiten verantwortlich? Abad und Kollegen (2003) stellten bei der Diskussion ihrer Befunde fest, dass der Effekt der Differenzierung für Testbatterien (z.B. der Wechsler-Intelligenztest), die auf kristalline Fähigkeiten fokussieren, stärker ist als bei Testbatterien, die fluide Fähigkeiten fokussieren. Als eine Erklärung hierfür führen sie an, dass Personen mit höherem Fähigkeitsniveau in der Regel meist eine bessere Schulbildung also Zugang zu besseren Investitionsmöglichkeiten haben. Wenn mit zunehmendem Niveau der Schulbildung Schüler mehr Wissen erwerben, das lösungsrelevant für Aufgaben zur Messung kognitiver (insbesondere kristalliner) Fähigkeiten ist, werden diese Aufgaben mit zunehmender Schulbildung weniger komplex. Die Reasoninganforderungen der Aufgaben nehmen ab, und somit werden auch die Ladungen der Aufgaben auf dem Faktor geringer, der die allgemeine kognitive Fähigkeit repräsentiert. Dies impliziert eine Differenzierung der kognitiven Fähigkeiten.

*Zusammenfassung.* Die referierten Studien weisen darauf hin, dass Personen mit hohem Fähigkeitsniveau eine differenziertere Faktorstruktur aufweisen als Personen mit einem niedrigen Fähigkeitsniveau. Von den aufgeführten sechs Studien (exkl. Spearman, 1927), die eine Differenzierung der kognitiven Fähigkeiten fanden, hatte nur die Testbatterie in der Studie von Carlstedt (2001) keine mathematischen Aufgaben enthalten. Da von den Autoren nichts Gegenteiliges berichtet wurde, lassen die Ergebnisse der fünf Studien vermuten, dass diese mathematikspezifischen Fähigkeitsfaktoren mit zunehmender allgemeiner kognitiver Fähigkeit differenzieren.

Erklärt wurde der Effekt der Faktordifferenzierung bzw. Faktorintegration durch Analogien zur Physik und Ökonomie (Spearman), die Automatisierung durch Fertigkeitserwerb sowie Investition in diese Fertigkeitserwerbsprozesse (Jensen) und als Effekt schulischen Lernens (Abad u.a.). Die Erklärung von Abad und Kollegen ist besonders interessant mit Blick auf mathematische Schülerleistung und mathematische Fähigkeiten: Ist die Differenzierung mathematischer Fähigkeiten eine Funktion der Schulform? Zumindest aus theoretischer Sicht können – wie im vorangegangenen Abschnitt ausgeführt wurde – verhaltensspezifische Unterschiede in der schulischen Lernumwelt zur Faktordifferenzierung führen. Eine empirische Antwort wird im Rahmen der zweiten Forschungsfrage zu Schulformunterschieden gegeben.

Insgesamt gesehen legen die Überlegungen zu Unterschieden der Lernumwelt in Abschnitt 4.3.1 und 4.3.2 nahe, schulischem Lernen eine Schlüsselrolle bei der Entwicklung und Differenzierung kognitiver Fähigkeiten zuzuschreiben (siehe auch Ceci, 1991). Deshalb beschäftigt sich das nächste Kapitel mit schulischem Lernen im Allgemeinen und Mathematikunterricht sowie schulformspezifischen Unterschieden des Mathematikunterrichts im Besonderen.

## 5 Mathematische Schülerleistung und schulisches Lernen

In der Unterrichtsforschung werden in jüngerer Zeit schulische Lernprozesse im Rahmen eines *Angebot-Nutzung-Modells* beschrieben (Baumert & Köller, 2000; Baumert u.a., 2004; Helmke, 2003; Schatz Koehler & Grouws, 1992). In diesem Modell wird der Schüler als der „aktive“ Konstrukteur seines Wissens und der Unterricht als ein „Angebot“ des Lehrers an die Schüler gesehen. Entscheidend für den Lernerfolg ist, dass der Lernende das Angebot nutzt, indem er die Lerninhalte im Unterricht aktiv mental verarbeitet. Unterricht bietet somit „nur“ eine Gelegenheitsstruktur für Lernprozesse.

Mit Blick auf mathematische Schülerleistung werden im Rahmen dieses Modells in den nächsten Abschnitten folgende Fragen erörtert: Wie kann man das Zusammenwirken von Angebot und Nutzung beschreiben (Abschnitt 5.1)? Welche individuellen kognitiven und motivationalen Schülermerkmale beeinflussen die Nutzung der Lernangebote und somit auch den Lernerfolg? Warum ist das so (Abschnitt 5.2)? Wie kann man die Angebotsstruktur „Mathematikunterricht“ differenziert beschreiben (Abschnitt 5.3)? Wie unterscheidet sich die Angebotsstruktur zwischen verschiedenen Schulformen (Abschnitt 5.4)?

Antworten auf diese Fragen stehen in einem engen Zusammenhang mit den Forschungsfragen. Die in Abschnitt 5.1 und 5.2 referierten Theorien und Befunde untermauern, mathematische Schülerleistung als kognitive Fähigkeiten zu konzeptualisieren (vgl. Forschungsfrage 1). Im Abschnitt 5.2.2 werden mit Blick auf die dritte Forschungsfrage die oftmals replizierten und somit sehr robusten Zusammenhänge zwischen motivationalen Schülermerkmalen und mathematischer Schülerleistung erläutert. Diese theoretische und empirisch fundierte Grundlage dient zur Ableitung von Hypothesen für die dritte Forschungsfrage zur Validität. Schulformspezifische Unterschiede lernförderlicher Aspekte des Mathematikunterrichts (Abschnitt 5.4, siehe auch Abschnitt 4.3) stützen Annahmen für die zweite Forschungsfrage, dass sich Schüler verschiedener Schulformen hinsichtlich dieser kognitiven Fähigkeiten im Niveau, aber auch hinsichtlich ihrer Heterogenität in den mathematischen Fähigkeiten unterscheiden können.

### 5.1 Modelle schulischen Lernens

In der Unterrichtsforschung werden mehrere Modelle des schulischen Lernens diskutiert (zusammenfassend z.B. Gruehn, 2000), die herangezogen werden können, um das Zusammenwirken von Angebot und Nutzung näher zu beleuchten. Prominente Vertreter sind das Modell von Carroll (1963, 1989), das Modell von Bloom (1976), das Modell von Walberg und Kollegen (Haertel, Walberg & Weinstein, 1983; Walberg, 1986; Wang, Haertel & Walberg, 1993) sowie die Modelle von Slavin (1994) und Creemers (1994). Da Carrolls Modell viele weitere Modelle inspirierte (z.B. Bloom, 1976; Creemers, 1994; Slavin, 1994) und Aspekte von Unterrichtsquantität, Unterrichtsqualität und individuellen Schülermerkmalen elegant über die aufgewendete Lernzeit verbindet, wird zunächst stellvertretend für Modelle schulischen Lernens darauf eingegangen. Anschließend werden zentrale Aspekte stichpunktartig skizziert, die den Modellen schulischen Lernens gemein sind, und es wird gezeigt, worin sie sich von Carrolls Modell unterscheiden.

*Carrolls Modell schulischen Lernens.* In Carrolls Modell (1963) des schulischen Lernens ist die Rolle der *aktiven, tatsächlich aufgewendeten Lernzeit* zentral (Abb. 12). Der Grad des Lernerfolgs eines Schülers für eine bestimmte Lernaufgabe (z.B. die Aneignung oder das Verstehen von mathematischen Begriffen oder das Erlernen mathematischer Fertigkeiten) ist eine Funktion des Verhält-

nisses von aufgewendeter Lernzeit und der benötigten Lernzeit. Je mehr die aufgewendete Lernzeit die benötigte Lernzeit überwiegt, desto mehr lernen Schüler hinzu. Welche Größen beeinflussen die aufgewendete und benötigte Lernzeit?

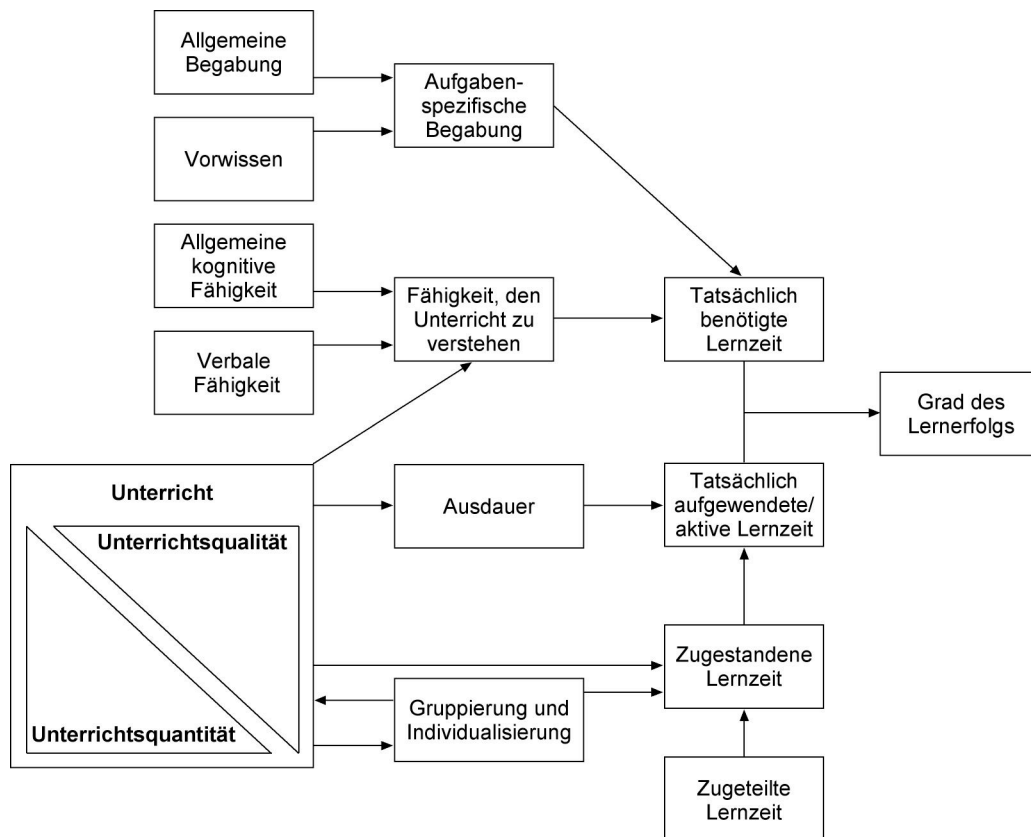


Abbildung 12: Carrolls Modell (1963) des schulischen Lernens in Anlehnung an Harnishfeger und Wiley (1977) und Gruehn (2000)

Die *benötigte Lernzeit* ist abhängig von der aufgabenspezifischen Begabung eines Schülers und seiner Fähigkeit, den Unterricht zu verstehen. Je höher die aufgabenspezifische Begabung ist, desto kürzer ist die Lernzeit für eine Aufgabe unter optimalen lernförderlichen Bedingungen (dem „idealen“ Unterricht). Die aufgabenspezifische Begabung ist dabei selbst wiederum eine Funktion der *allgemeinen Begabung* (als stellvertretender Begriff für eine Vielzahl leistungsrelevanter Schülereigenschaften und -fähigkeiten) und des aufgabenrelevanten Vorwissens. Die *Fähigkeit, den Unterricht zu verstehen*, wird von der allgemeinen kognitiven Fähigkeit und der verbalen Fähigkeit beeinflusst und interagiert mit der Unterrichtsqualität. Je mehr Schüler im Unterricht Begriffe und Beziehungen aus dem Unterricht selbst erschließen müssen und je weniger der Lehrer diese Beziehungen für alle Schüler klar herausstellt, desto stärker ist der Einfluss der allgemeinen kognitiven Fähigkeit. Die verbale Fähigkeit spielt dann eine Rolle, wenn die sprachlichen Anforderungen im Unterricht sehr hoch sind.

Die *aktive (aufgewendete) Lernzeit* ist abhängig von der gezeigten Ausdauer und der zugestandenen Lernzeit. Die *Ausdauer* oder Lernmotivation eines Schülers konzipiert Carroll als die Zeit, die ein Schüler bereit ist, aktiv und aufmerksam etwas zu lernen. Die Ausdauer kann dabei sowohl durch extrinsische (z.B. Noten) als auch intrinsische (z.B. durch interessante oder verblüffende mathe-

matische Probleme, z.B. Mitchell, 1993) Anreize im Unterricht beeinflusst werden. Die *zugestandene Lernzeit* ist die Zeit, die dem Schüler zum Lernen zur Verfügung steht. Diese wird einerseits durch Gruppierungs- und Individualisierungsmaßnahmen bestimmt, die Lerntempo, Lerninhalt oder Stoffmenge an die Schülereigenschaften anpassen (Harnishfeger & Wiley, 1977). Andererseits ist die zugestandene Lernzeit eine Funktion der in der Stundentafel festgelegten Unterrichtsstunden (*zugeteilte Lernzeit*) und der vom Lehrer für lehrstoffrelevante Aktivitäten genutzten Lernzeit (Gruehn, 2000). Ein weiterer Aspekt der zugeordneten Lernzeit sind die Hausaufgaben, die eine Lehrkraft aufgibt (Trautwein & Köller, 2003).

Der Unterricht nimmt in Carrolls Modell eine Schlüsselrolle für den Lernerfolg ein. Ziel des Lehrers sollte es sein, die zur Verfügung stehende Zeit möglichst effektiv zum Schaffen von Lerngelegenheiten zu verwenden und die aktive Auseinandersetzung der Schüler mit dem Lernstoff zu stimulieren (z.B. in Form selbstregulierter Lernprozesse). Daher stehen dem Lehrer mehrere „Hebel“ zur Verfügung, dieses Ziel zu erreichen: Er wählt die Unterrichtsmaterialien aus, beeinflusst die zugestandene Lernzeit, die Ausdauer der Schüler (z.B. durch motivationale Anreize) und die Anforderungen an die Fähigkeit der Schüler, den Unterricht zu verstehen.

Zusammengefasst sieht das Zusammenwirken von Angebot und Nutzung in Carrolls Modell folgendermaßen aus. Die Aktivitäten des Lehrers und das Unterrichtsgeschehen sind zunächst zum Lernerfolg distal. Sie bestimmen die Anforderungen an die Fähigkeit der Schüler, den Unterricht zu verstehen, die zugestandene Lernzeit und die Lerninhalte. Nur wenn ein Schüler dieses Angebot tatsächlich nutzt, indem er sich aktiv mit den Lerninhalten auseinandersetzt und hierfür Lernzeit investiert, ist das Lernen erfolgreich. Die Aktivität des Schülers wird dabei durch kognitive Fähigkeiten und Vorwissen sowie motivationale Variablen reguliert. Der Ertrag der Lernprozesse ist umso größer, je positiver das Verhältnis von aufgewendeter aktiver Lernzeit des Schülers zur benötigten Lernzeit ausfällt.

*Gemeinsamkeiten und Unterschiede.* Wie eingangs dargestellt, versuchen auch andere Modelle schulischen Lernens das Zusammenwirken von Angebot und Nutzung zu erklären. Worin unterscheiden sich diese Modelle von Carrolls Modell, was ist ihnen gemein?

Gemein ist allen Modellen schulischen Lernens die Annahme, dass Quantität und Qualität des Unterrichts *sowie* kognitive und motivationale Schülereigenschaften die Entwicklung der Schülerleistung in einem bestimmten Fach beeinflussen (siehe auch Helmke & Weinert, 1997).

Unterschiede bestehen in folgenden Aspekten (zusammenfassend z.B. Gruehn, 2000):

- Bloom (1976) stellt im Vergleich zu Carroll die Lernzeit nicht so sehr in den Mittelpunkt. Das Modell konzentriert sich auf den sequenziellen und kumulativen Charakter von schulischen Lernprozessen: Das Lösen einer bestimmten Lernaufgabe geht mit motivationalen Veränderungen und Wissenserwerb einher. Diese motivationalen und kognitiven Veränderungen wirken sich im Gegenzug auf das weitere Lernen aus. Somit bezieht Bloom im Gegensatz zu Carroll (1963, S. 724) auch motivationale Lernergebnisse in das Modell mit ein.
- Creemers (1994) und Slavin (1994) differenzieren das Konstrukt Unterrichtsqualität stärker als Carroll und integrieren dabei die Befundlage aus der Unterrichtsforschung.

- Walberg und Kollegen beziehen zusätzlich zu Schülermerkmalen und Unterricht weitere *direkte* leistungsrelevante Einflussfaktoren in ihr Modell mit ein, wie zum Beispiel den familiären Hintergrund.

Vor dem Hintergrund dieser Unterschiede können einige Aspekte im Modell von Carroll noch weiter qualifiziert werden:

- Carroll behandelt in seinem Modell die Lernmotivation nur sehr undifferenziert als Determinanten der „Ausdauer“ und klammert im Gegensatz zu Bloom motivationale Ergebnisse des Lernens aus. Daher werden in Abschnitt 5.2.2 verschiedene Aspekte der Motivation unterschieden und das Zusammenwirken von motivationalen Facetten und Lernprozessen betrachtet.
- Das Modell von Carroll ist ein allgemeines Modell, das nicht für ein bestimmtes Unterrichtsfach zugeschnitten ist. Zudem ist in seinem Modell das Konstrukt „Unterrichtsqualität“ im Vergleich zu den Modellen von Creemers (1994) und Slavin (1994) weniger differenziert (Gruehn, 2000). Damit stellt sich die Frage, welche Anforderungen im Unterricht gestellt werden und welche Aspekte des Mathematikunterrichts das Lernen beeinflussen (siehe Abschnitt 5.3 und 5.4).
- Wie kognitive und motivationale Schülermerkmale zum „Wissenserwerb“ führen, beschreibt Carroll nur sehr knapp. Dies ist Gegenstand des nächsten Abschnitts.

## 5.2 Schülermerkmale und schulisches Lernen

### 5.2.1 Kognitive Fähigkeiten, Wissen und Wissenserwerbsprozesse

Die nachfolgenden Theorien und Überlegungen zu Wissenserwerbsprozessen und der besonderen Bedeutung von Wissen bei der Informationsverarbeitung wurden nicht mathematikspezifisch formuliert. Sie werden aber von einigen Autoren auf mathematisches Problemlösen und Erwerbsprozesse von mathematischem Wissen generalisiert (z.B. Geary, 2005; Weinert, 1996). Die dargestellten Theorien und Befunde werden auch deshalb ausgewählt, da sie den Zusammenhang zwischen *interindividuellen Unterschieden* in kognitiven Fähigkeiten mit Informationsverarbeitungs-, Wissens- und Fertigkeitserwerbsprozessen näher erläutern (für einen umfassenden Überblick zu diesem Thema siehe auch Ackerman, Kyllonen & Roberts, 1999; Ackerman, Sternberg & Glaser, 1989).

*Die Investmenttheorie von Cattell (1987).* Die Investmenttheorie beschreibt, wie die fluide Fähigkeit<sup>11</sup> zum Erwerb kristalliner Fähigkeiten (z.B. die generelle mathematische Fähigkeit im Gf-Gc-Modell; vgl. Abschnitt 4.2) beiträgt. Um die Funktion fluider Fähigkeit bei Wissenserwerbsprozessen zu beschreiben, unterscheidet Cattell Prozesse, die Einsicht in komplexe Relationen erfordern, von solchen Prozessen, in denen dies nicht notwendig ist. Bei Letzteren hängt die Lernrate (in Form des Lernzugewinns) vornehmlich von der Lernmotivation, von einfachen Gedächtnisfunktionen oder der Häufigkeit positiver Verstärkung ab. Bei Prozessen, die Einsicht in komplexe Relationen erfordern (z.B. zu lernen, wie man arithmetische Textaufgaben löst), ist jedoch die Lernrate in hohem Maße durch das Niveau der fluiden Fähigkeit bestimmt.

---

<sup>11</sup> Nach Gustafsson (1984) kann die fluide Fähigkeit als nahezu äquivalent zur allgemeinen kognitiven Fähigkeit betrachtet werden.

Cattell nimmt an, dass die fluide Fähigkeit in alle komplexen Lernaufgaben *investiert* wird und zum Erwerb kristalliner Fähigkeiten führt. Das erreichte Niveau kristalliner Fähigkeiten ist das Ergebnis kumulativer Lernprozesse durch die vielfache Investition der fluiden Fähigkeit. Cattell geht davon aus, dass die fluide Fähigkeit und der Erwerb kristalliner Fähigkeiten positiv miteinander korrelieren, also Personen mit höherem Niveau fluider Fähigkeit auch ein höheres Niveau bei kristallinen Fähigkeiten erreichen.

Allerdings konkretisiert Cattell in seiner Investmenttheorie nicht genauer, wie der Investitionsprozess beim Erwerb kristalliner Fähigkeiten abläuft. Zur Beschreibung dieses Prozesses ist ein Phasenmodell des Fertigkeitserwerbs hilfreich.

*Fertigkeitserwerb.* Ackerman (1987, 1988, 1989; Ackerman & Kanfer 1993) geht in Einklang mit Theorien aus der Forschung zur Informationsverarbeitung (z.B. Anderson, 1993) davon aus, dass der Fertigkeitserwerb in drei Phasen abläuft. In der *kognitiven Phase* sind Gedächtnis- und Reasoningprozesse notwendig. Die Bearbeitung von Aufgaben in dieser Phase ist langsam und fehleranfällig, die gesamte Aufmerksamkeit eines Lerners ist notwendig, um die Aufgabe zu verstehen und auszuführen. In dieser Phase werden hohe Anforderungen an die Aufmerksamkeitskontrolle (Ackerman, 1988) und das Arbeitsgedächtnis gestellt (Anderson, 1993).

Die Leistungen in dieser Phase korrelieren am höchsten mit der allgemeinen kognitiven Fähigkeit. Hierfür spricht neben der empirischen Unterstützung des Modells von Ackerman (1988), dass die allgemeine kognitive Fähigkeit hoch mit der Arbeitsgedächtniskapazität interkorreliert ist (Colom, Rebollo, Palacios, Juan-Espinosa & Kyllonen, 2004; Kane, Hambrick & Conway, 2005; Kyllonen & Christal, 1990; Oberauer, Schulze, Wilhelm & Süß, 2005) und die Arbeitsgedächtniskapazität in dieser Phase (sensu Anderson) eine entscheidende leistungslimitierende Ressource darstellt.

Sobald ein Lerner eine adäquate mentale Repräsentation der Aufgabe gebildet hat, geht er über in die *assoziative Phase*. Während der assoziativen Phase fügen Lernende die Sequenzen kognitiver und psychomotorischer Prozesse, die zur Aufgabenbearbeitung notwendig sind, zusammen (Kompilierung). Gleichzeitig wird die assoziative Verbindung zwischen Stimulus (als Bedingungen, die die Handlungen erlauben) und Response (das sind die konkreten Handlungen selbst) im Sinne einer Produktionsregel fester. Bei konsistenten Aufgabenanforderungen erhöhen sich so mit zunehmender Übung die Bearbeitungsgeschwindigkeit und die Akkuratheit der Bearbeitung. Gleichsam reduzieren sich die Anforderungen an die Aufmerksamkeitskontrolle und an das Arbeitsgedächtnis.

Die Leistungen in der assoziativen Phase korrelieren am höchsten mit der Informationsverarbeitungsgeschwindigkeit, die Ackerman (1989, S. 179) als „perceptual speed ability“ bezeichnet. Tests, die diese Fähigkeit messen, erfordern häufig, dass innerhalb einer Inhaltsdomäne (verbal, numerisch oder figural-räumlich) Stimuli miteinander verglichen und Unterschiede festgestellt werden. In den Termini von Informationsverarbeitungsprozessen betrifft es also die Fähigkeit zur Generierung einfacher Bedingung-Handlungssequenzen bzw. Stimulus-Response-Assoziationen. Bedingungen sind Gleichheit oder Ungleichheit der Stimuli, die mit den Handlungen „Nonresponse“ oder Indizieren der Ungleichheit assoziiert sind. Höhere Punktleistungen bei diesen Tests korrespondieren mit einer stärker ausgeprägten Fähigkeit, diese einfachen Assoziationen zu generieren. Dies entspricht damit den Anforderungen in der assoziativen Phase.

In der *autonomen Phase* ist die Aufgabenbearbeitung weitestgehend automatisiert und kann so ohne große Aufmerksamkeitsanforderungen ausgeführt werden. Während dieser Phase wird die zu lernende Fertigkeit internalisiert, sodass bei Präsentation des Stimulus die Antwort ohne bewusste Kontrolle des Lernenden abgerufen werden kann. Durch langandauerndes Üben bei konsistenten Aufgabenanforderungen wird die Aufgabenbearbeitung schnell und akkurat, selbst wenn man gleichzeitig eine zweite Aufgabe ausführen muss (dies war in der assoziativen Phase noch nicht möglich).

Die Leistungen in dieser Phase korrelieren am höchsten mit psychomotorischen, im weiteren Sinne nicht kognitiven Fähigkeiten. Interindividuelle Unterschiede bestehen hier in der Geschwindigkeit und Genauigkeit motorischer „Antworten“. Es ist klar, dass diese Korrelation mit der Lernleistung nur bei Aufgaben besteht, die motorische Anforderungen stellen (vgl. Ackerman, 1988, S. 311).

Bei Aufgaben mit konsistenten Anforderungen kann also der Ablauf des Fertigkeitserwerbs in drei Phasen beschrieben werden. Zentrales Ergebnis ist, dass mit zunehmender Übung Personen schneller und genauer beim Bearbeiten von Aufgaben werden. Die Lösung von anfangs neuen Aufgaben wird im Laufe der Übung mehr und mehr prozeduralisiert und damit als Produktionsregel Bestandteil der (prozeduralen) Wissensbasis. Jedoch ist eine so beschriebene Prozeduralisierung bei inkonsistenten Aufgabenanforderungen nicht möglich. Wenn die Inkonsistenz der Aufgabenkomponenten dominiert, stoppt die Prozeduralisierung in der kognitiven Phase. Somit sind immer noch Gedächtnis- und Reasoningprozesse notwendig, und es besteht eine relativ hohe Arbeitsgedächtnisbelastung. Wenn Aufgaben konsistente und inkonsistente Elemente enthalten, hängt der Grad der Prozeduralisierung vom Verhältnis konsistenter und inkonsistenter Elemente ab (Ackerman, 1989).

Eine wichtige Feststellung ist, dass anfangs neue Aufgaben auch als Probleme betrachtet werden können. Folgt man den Annahmen des Phasenmodells des Fertigkeitserwerbs, so führt wiederholtes Lösen von Problemen mit konsistenter Struktur zur Prozeduralisierung der lösungsrelevanten Produktionsregeln, die dann Bestandteil der prozeduralen Wissensbasis werden: Problemlösen ist nach Anderson (1993) ein Lernprozess. Wichtig ist hierbei festzuhalten, dass neben der allgemeinen kognitiven Fähigkeit (Süß, 1996) bereichsspezifisches Vorwissen einer der besten Prädiktoren der Problemlöseleistung ist (zusammenfassend z.B. Spada & Wichmann, 1996).

*Zur Rolle des Vorwissens.* Weshalb hat Vorwissen (im weiteren Sinne) positive Auswirkungen auf die Problemlöseleistung? Nach Renkl (1996) erleichtert das Vorwissen mehrere Prozesse der Informationsverarbeitung. Bei der *Informationsselektion* ermöglicht das Vorwissen, die Aufmerksamkeit auf lösungsrelevante Aspekte zu lenken. Bei der *Encodierung der Information* kann die aufgenommene Information eher mit dem Vorwissen in Beziehung gesetzt werden. Zudem können mithilfe des Vorwissens größere Chunks gebildet werden. Damit wird die Arbeitsgedächtniskapazität weniger stark ausgelastet, und es sind mehr Ressourcen für weitere Informationsverarbeitungsprozesse verfügbar. Bei der *Informationsspeicherung* können neue Informationen an bereits bestehende Schemata angebunden, in das bestehende, geordnete Wissensnetz integriert und mit „Sinn“ versehen werden. Wissen erleichtert auch den *Informationsabruf* aus dem Langzeitgedächtnis. Besteht bereits eine reichhaltige Wissensbasis, so sind die einzelnen Wissenseinheiten vielfältig miteinander verknüpft. Geht man von einem assoziativen Netzwerk aus, in dem sich die einzelnen Wissenselemente wechselseitig aktivieren können, so stehen somit vielfältige Zugangswege zu diesen Wissenseinheiten offen. Damit können auch Informationen, die nicht direkt aus dem Gedächtnis abrufbar sind, eher rekonstruiert werden. Wurde zudem das Wissen bereits in vielfältigen Situatio-



nen eingesetzt, so sind die Anwendungsbedingungen für einzelne (Problemlöse-)Schemata eher bekannt. Insbesondere durch Automatisieren von mathematischen Prozeduren (in Form von prozeduralisiertem Wissen) können Schüler die Belastung ihres Arbeitsgedächtnisses zusätzlich verringern und diese freigewordenen Ressourcen für höhere (meta-)kognitive Prozesse (z.B. das Überwachen des eigenen Problemlöseverhaltens) verwenden (siehe auch Mayer, 1994; Sweller, 1989).

*Zusammenfassung.* Das Wirken kognitiver Schülermerkmale auf die Lernleistung in Mathematik kann man knapp wie folgt zusammenfassen (vgl. Helmke & Weinert, 1997): Je stärker die allgemeine kognitive Fähigkeit (bzw. die hierzu als äquivalent angenommene fluide Fähigkeit) von Schülern ausgeprägt ist, desto besser können sie sich auf neue (komplexe) mathematische Probleme einstellen, effektive Problemlösestrategien entwickeln und lösungsrelevante Regeln erkennen. Dies entspricht den Annahmen der Investmenttheorie von Cattell wie auch den Annahmen im Modell des Fertigkeitserwerbs. In beiden Theorien ist die allgemeine kognitive Fähigkeit eine Determinante der interindividuellen Unterschiede in den Lernraten bzw. in Anlehnung an die Theorie von Ackerman den Unterschieden der Lernraten zu Beginn der Fertigungsprozeduralisierung. So erwerben Schüler mit stärker ausgeprägter allgemeiner kognitiver Fähigkeit unter vergleichbaren Zeit- und Instruktionsbedingungen mehr Mathematikwissen (im weiteren Sinne). Folgt man der Argumentation von Abschnitt 4.1 zum Zusammenhang von Schülerleistung, Wissen und Fähigkeiten, können diese interindividuellen Unterschiede im Mathematikwissen beispielsweise durch interindividuelle Unterschiede in der generellen mathematischen Fähigkeit (vgl. Abschnitt 4.2, Gf-Gc-Modell) repräsentiert werden.

Das mathematikspezifische Vorwissen selbst fördert wiederum den Erwerb neuen mathematischen Wissens, da es mehrere lernrelevante Aspekte der Informationsverarbeitung sowie späteres mathematisches Problemlösen erleichtert, wie die Überlegungen von Renkl (1996) nahe legen. Wenn nun ein mathematisches Problem neuartige (bzw. unbekannte) und bekannte Aufgabenelemente enthält, können zur Lösung dieses Problems die allgemeine kognitive Fähigkeit und mathematisches Wissen beitragen (siehe auch Boekaerts, 1996, S. 105). Interindividuelle Unterschiede in der allgemeinen kognitiven Fähigkeit und im mathematischen Wissen sind somit leistungsrelevante Determinanten beim Lösen mathematischer Probleme.

Neben kognitiven Determinanten ist eine wesentliche Einflussgröße für das Lösen von Problemen auch wie ausdauernd und motiviert Schüler sich dem Problemlöseprozess widmen und versuchen, eine Lösung zu finden. Der nächste Abschnitt behandelt daher motivationale Aspekte beim Lernen in Mathematik.

### 5.2.2 Lernmotivation

Motivationale Aspekte im schulischen Kontext werden im angloamerikanischen Sprachraum unter dem Begriff *achievement motivation* zusammengefasst. Rheinberg (1996) spricht hierbei von „Lernmotivation“. Die Lernmotivation von Jugendlichen beeinflusst, welche Lernsituationen sie aufsuchen, wie sehr sie sich anstrengen, wie lange sie sich mit Problemen in diesen Lernsituationen beschäftigen und damit natürlich auch wie erfolgreich sie beim Problemlösen sind (Eccles & Wigfield, 2002; Heckhausen, 1989; Pintrich, 2003; Wigfield & Eccles, 2000). Die positive Wirkung der Lernmotivation wird (zumindest teilweise) über die eingesetzten selbstregulativen Lernstrategien der Schüler mediiert (Bruinsma, 2004; Covington, 2000; Pekrun & Schiefele, 1997; Schiefele & Pekrun, 1997). Mit Covington (2000) kann man die Befunde folgendermaßen zusammenfassen: Je positiver die Komponenten der Lernmotivation ausgeprägt sind,

- desto tiefer und elaborierter wird der Lernstoff verarbeitet,
- desto mehr wird der eigene Lernfortschritt überwacht,
- desto mehr werden Strategien zur Organisation des Lernstoffs (z.B. Zusammenfassen oder in eigenen Worten paraphrasieren) herangezogen und
- desto mehr werden gelegentlich ausbleibende Lernfortschritte so attribuiert, dass dies keine negativen Auswirkungen auf das Begabungsselbstkonzept und die eigene Kontrollwahrnehmung hat.

Eine der traditionellen Perspektiven im Rahmen der Forschung zur Lernmotivation sind „Erwartung-Wert-Theorien“ der Lernmotivation. Diese Theorien ermöglichen eine elegante Klassifikation (Pintrich, 2003) einer Vielzahl motivationaler Konstrukte (für Übersichten siehe z.B. Pintrich, 2003; Snow, Corno & Jackson, 1996), die in der pädagogischen Psychologie diskutiert werden. Nachfolgend werden stellvertretend zwei ausgewählte motivationale Konstrukte als typische Vertreter von Erwartungs- und Wertkomponenten etwas detaillierter dargestellt.

Zu den *Erwartungskomponenten* kann man mit Pintrich (2003) und Eccles und Wigfield (2002) Kontrollüberzeugungen, Ergebnis- und Selbstwirksamkeitserwartungen und mit Eccles und Wigfield (2002) auch das (fachspezifische) Selbstkonzept zählen.

Während Kontrollüberzeugungen und die Konstrukte aus der Selbstwirksamkeitstheorie Erwartungen über zukünftige Ereignisse betreffen, sind *Selbstkonzepte* selbstbezogene Kognitionen über die eigene gegenwärtige Fähigkeit. Selbstkonzepte betreffen nach Helmke und Weinert (1997) die eigenen leistungsbezogenen Kompetenzen, Ressourcen und Kapazitäten, wie auch das Vertrauen in die eigene Leistungsfähigkeit (z.B. in Mathematik). Selbstkonzepte können dabei auch nach dem Grad ihrer Fachspezifität unterschieden werden (z.B. allgemeines akademisches Selbstkonzept vs. mathematisches oder verbales Selbstkonzept).

Insgesamt gesehen betreffen Erwartungskomponenten die Antworten, die sich Schüler auf die Frage „Kann ich diese Aufgabe bearbeiten?“ („Can I do this task?“; Pintrich, 2003, S. 105) geben.

Die *Wertkomponente* hingegen umfasst, wie Schüler die Frage „Warum beschäftige ich mich mit dieser Aufgabe?“ beantworten („Why am I doing this task?“; Pintrich, 2003, S. 109). Zu den Wertkomponenten gehören der mit einer Handlung verbundene Nutzen und die entstehenden Kosten sowie der persönliche Wert, der mit dem Erreichen eines bestimmten Handlungsergebnisses assoziiert ist. Weiterhin zählt zu den Wertkomponenten der intrinsische Wert der Handlung selbst (Eccles & Wigfield, 2002; Wigfield & Eccles, 2000).

Eng zusammenhängend mit dem intrinsischen Wert einer Handlung (z.B. mathematischem Problemlösen) ist das fachspezifische Interesse (z.B. in Mathematik). Krapp (1998, 2002) betrachtet Interessen aus dem Blickwinkel einer Person-Objekt-Beziehung. Fachspezifisches Interesse stellt dabei ein stabiles Persönlichkeitsmerkmal dar und zeichnet sich vornehmlich durch persönliche Valenz oder Bedeutsamkeit (z.B. „Mathematik ist mir wichtig“) und positiv-affektive Valenz (z.B. „Mathematik macht mir Spaß“) eines Objekts (z.B. Mathematik) und den damit verbundenen Handlungsmöglichkeiten (mathematisches Problemlösen) aus.

Aus den Modellen schulischen Lernens kann gefolgert werden, dass die mathematikspezifische Lernmotivation und mathematische Schülerleistung positiv kovariieren. Mit Blick auf das mathematische Selbstkonzept wird diese Annahme von zahlreichen querschnittlichen Studien unterstützt (z.B. Köller, Schnabel & Baumert, 2000, Tab. 1; Lüdtke, Köller, Artelt, Stanat & Baumert, 2002). Eine etwas ältere Zusammenfassung dieser Studien bietet die Meta-Analyse von Hansford und Hattie (1982, S. 134, Tab. 8). Sie fanden eine durchschnittliche Korrelation von .20 zwischen verschiedenen Maßen selbstbezogener Kognitionen (z.B. das Selbstkonzept) und tatsächlicher Mathematikleistung. Eine Korrelation in ähnlicher Höhe ( $r = .23$ ) wird in einer Meta-Analyse jüngerer Datums von Ma und Kishor (1997, S. 99, Tab. 1) berichtet. Die neueste Arbeit stammt von Möller und Köller (2004). In ihrem Übersichtsartikel berichten die beiden Autoren eine Median-Korrelation (berechnet über 34 Studien) von .47 zwischen Leistungsindikatoren (Noten und Leistungstests) und dem mathematischen Selbstkonzept. In einer Meta-Analyse längsschnittlicher Studien fanden Valentine, DuBois und Cooper (2004) Evidenz dafür, dass sich ein positiveres fachspezifisches Selbstkonzept (unter Kontrolle zur zeitgleich gemessenen Schülerleistung) leicht positiv auf die spätere Schülerleistung auswirkt (für eine Einzelstudie, die diese Hypothese stützt, siehe z.B. Marsh & Yeung, 1997).

Ein ähnliches Bild wie für das mathematische Selbstkonzept zeigt sich auch für den Zusammenhang zwischen mathematischem Interesse und mathematischer Schülerleistung. In einer Meta-Analyse von überwiegend querschnittlichen Studien fanden Schiefele, Krapp und Schreyer (1993, Tab. 3) eine durchschnittliche Korrelation von .28 zwischen mathematischem Interesse und mathematischer Schülerleistung. In einer längsschnittlichen Analyse konnten Köller, Baumert und Schnabel (2001, Abb. 2) zeigen, dass mathematisches Interesse von Gymnasiasten in der 10. Klasse einen positiven Effekt auf die mathematische Schülerleistung in der 12. Klasse hat (auch wenn statistisch für leistungsrelevante Variablen wie frühere mathematische Schülerleistung und Kursniveau kontrolliert wird). In der Studie von Marsh und Kollegen (Marsh, Trautwein, Lüdtke, Köller & Baumert, 2005, Abb. 2.3) hatte in zwei unabhängigen Studien mathematisches Interesse (ohne Berücksichtigung anderer motivationaler Konstrukte) in der 7. Klasse einen leicht positiven Effekt auf mathematische Schülerleistung in der 8. Klasse (bei Kontrolle der Schülerleistung in der 7. Klasse).

Mit Blick auf das Fach Mathematik kann zusammenfassend festgehalten werden, dass sich mathematische Lernmotivation aus verschiedenen Aspekten zusammensetzt, die das Lernen und die Leistung in Mathematik beeinflussen. Der häufig replizierte Befund einer (querschnittlich gefundenen) positiven Kovariation zwischen mathematischer Lernmotivation und mathematischer Schülerleistung ist wohl das Ergebnis dreier Mechanismen:

- Die Lernmotivation beeinflusst die Anstrengung und Persistenz beim Problemlösen.
- Die Lernmotivation wirkt sich auf die Auswahl der eingesetzten selbstregulativen Lernstrategien aus.
- Die erzielte Leistung selbst wiederum hat (z.B. vermittelt über das eigene Kompetenzerleben, siehe hierzu Abschnitt 9.4 sowie Wigfield & Eccles, 2000) einen positiven Effekt auf die Lernmotivation. Diese Annahme wird von mehreren empirischen Studien bestätigt (z.B. Helmke & van Aken, 1995; Köller, Baumert u.a., 2001; Marsh u.a., 2005; Marsh & Yeung, 1997).

Unabhängig vom angenommenen Wirkmechanismus ist interessant, dass bei den referierten Arbeiten zum Zusammenhang von Lernmotivation und mathematischer Schülerleistung verschiedene Aspekte der Lernmotivation sehr differenziert und theoretisch elaboriert betrachtet werden. Hingegen wird aber in diesen Arbeiten weit weniger theoretisches Augenmerk darauf gelenkt, wie die kognitiven Fähigkeiten mathematischer Schülerleistung differenziert werden können. Dies hat wichtige Implikationen für die dritte Forschungsfrage zur externen Validität, wie in Kapitel 6 noch ausgeführt wird.

Resümierend ist festzustellen, dass die individuellen kognitiven und motivationalen Schülermerkmale wichtige Determinanten des Lernens in Mathematik sind. Wirkmechanismen wurden im Rahmen der Investmenttheorie, anhand des Phasenmodells zum Fertigkeitserwerb, anhand der Rolle des Vorwissens bei der Informationsverarbeitung und anhand von Komponenten der Lernmotivation im Rahmen von „Erwartung-Wert-Theorien“ diskutiert.

Entscheidend ist, dass diese Wirkmechanismen natürlich nur greifen können, wenn ein konkretes Lernangebot vorliegt. Für das Lernangebot hat die Schule für das Fach Mathematik nahezu eine Monopolstellung inne (Köller & Baumert, 2002). Daher ist es eine interessante Frage, wie Lernangebote im Mathematikunterricht gestaltet sind und wie der Unterricht dazu beitragen kann, dass Schüler die Fähigkeit zum mathematischen Problemlösen trainieren und mathematisches Wissen erwerben. Antworten darauf werden anhand von Überlegungen und der Befundlage aus der Unterrichtsforschung gegeben.

### **5.3 Mathematikunterricht**

Um die leistungsförderlichen Aspekte des Mathematikunterrichts (das Angebot) differenziert zu beschreiben, ist eine vierstufige Unterteilung des Curriculums hilfreich (Floden, 2002; Köller & Baumert, 2002). Die vier Stufen sind das intendierte Curriculum (Lehrpläne und Prüfungsvorschriften), das potenzielle Curriculum (von den Kultusministerien zugelassene Lehrbücher), das implementierte Curriculum (tatsächlich behandelter Stoff) und das erreichte Curriculum (tatsächliche Schülerleistungen).

Lernförderliche Aspekte des Mathematikunterrichts (als implementiertes Curriculum) können vor der normativen Folie des intendierten Curriculums beleuchtet werden. Wie in Abschnitt 2.1 bereits dargestellt, sollen im Rahmen deutscher Mathematiklehrpläne mathematisches Verständnis und Problemlösefähigkeiten, aber auch mathematische Standardverfahren und Lösungsroutinen (Kunter, 2005) vermittelt werden. Die Lerninhalte sind dabei in erster Linie nach mathematischen Stoffgebieten gegliedert. Jedoch steuern die Lehrpläne nur mittelbar das Unterrichtsgeschehen, da Lehrkräfte diese meist nur für bestimmte Anlässe wie Berufsantritt, Elternabende und Visitationen der Schulaufsicht heranziehen. In der Regel greifen sie auf lehrplankonforme, zugelassene Schulbücher (potenzielles Curriculum) zurück (Avenarius u.a., 2003).

Um die Lehrpläne im Unterricht zu implementieren, treffen Mathematiklehrkräfte in der Unterrichtsplanung eine Reihe von Entscheidungen im Vorfeld des Unterrichts (Schrader, 1997), zum Beispiel durch die Aufgabenauswahl aus den Lehrbüchern. Gleichsam schaffen sie während des Unterrichts Opportunitäten (siehe auch Bromme, 1997) für Lern- und Problemlöseprozesse, aber auch für das Lernen von Definitionen und für das Üben von mathematischen Operationen.

*Basisdimensionen der Unterrichtsqualität.* In der Unterrichtsforschung (für eine Zusammenfassung siehe Brophy & Good, 1986; Einsiedler, 1997) konnten mehrere *fächerübergreifende Basisdimensionen* der Unterrichtsqualität identifiziert werden, die das Erreichen der curricularen Lernziele unterstützen (Baumert & Köller, 2000; Baumert u.a., 2004; Brophy, 2000; Gruehn, 2000; Helmke, 2003). Nach Baumert und Kollegen können folgende Aspekte unterschieden werden:

- Eine störungspräventive Unterrichtsführung und effektive Behandlung von kritischen Ereignissen ermöglichen eine erhöhte Aufmerksamkeit der Schüler für die Lerninhalte (effektive Klassenführung, vgl. Klieme, Schümer u.a., 2001).
- Das Unterrichtstempo, mit dem der Stoff durchgenommen wird, wie auch das Interaktions-tempo sollten so sein, dass die Schüler nicht überfordert sind und Zeit zum Nachdenken haben.
- Der Lernstoff und die Aufgabenstellungen sollten klar, verständlich und strukturiert dargeboten werden.
- Der Unterricht sollte die Schüler motivieren und für die Lerninhalte interessieren (vgl. Bromme, 1997; Mitchell, 1993), und die Lehrer-Schüler-Beziehung sollte von wechselseitigem Verständnis geprägt sein (sozio-affektive Qualität).
- Lehrplan, Lerngelegenheiten im Unterricht und Leistungsbeurteilung sollten konsistent sein. Grundsätzlich gilt, dass die Lernaufgaben und Lehrmaterialien im Unterricht, die die Schüler bearbeiten, maßgeblich die Lernergebnisse beeinflussen. Im Unterricht erworbenes Wissen und erworbene Fertigkeiten können in der Regel umso leichter zur Problemlösung der Testaufgaben, die zur Überprüfung der Lernerfolge eingesetzt werden, transferiert werden, je ähnlicher sich Übungsaufgaben und die Testaufgaben sind (Mayer, 2004; Mayer & Wittrock, 1996). So konnten beispielsweise Muthén, Kao und Burstein (1991) und Klieme und Baumert (2001) zeigen, dass Schüler die Mathematikaufgaben tendenziell besser lösen konnten, die sie im Unterricht behandelt hatten.
- Viele Forscher (Bloom, 1976; Carroll, 1963; Corno u.a., 2002; Corno & Snow, 1986; Helmke & Schrader, 1987; Slavin, 1994; Snow, 1989; Snow & Swanson, 1992) gehen davon aus, dass der optimale Lerneffekt erzielt wird, wenn die Anforderungen der Aufgaben für die Schüler genau bei oder etwas über ihren Fähigkeiten (im weiteren Sinne) liegen. Die im Unterricht gestellten Aufgaben sollten also weitestgehend an das Fähigkeitsniveau eines Schülers adaptiert sein.

Zusammenfassend zeichnet sich hohe Unterrichtsqualität also zum einen durch die Güte der instruktionalen Maßnahmen und zum anderen durch eine möglichst effektive Nutzung der zur Verfügung stehenden Unterrichtszeit (Unterrichtsquantität) aus. Die Basisdimensionen der Unterrichtsqualität stellen lernförderliche Aspekte des Unterrichts dar, die aber mathematikspezifische Aspekte des Lernens außer Acht lassen (vgl. Ball, Lubienski & Mewborn, 2001; Baumert u.a., 2004; Mayer, 2004). Damit stellt sich die Frage, welche mathematikspezifischen Aspekte des Unterrichts den Lernerfolg beeinflussen.

*Mathematikspezifische Aspekte der Unterrichtsqualität.* Zu den effektiven instruktionalen Maßnahmen, die das Üben und Sichern mathematischer Standardverfahren und Operationen ermöglichen, können Formen des *drill and practice* gerechnet werden. Es ist eine lernpsychologische Tatsache, dass das oftmalige, wiederholte Bearbeiten der gleichen oder ähnlichen Aufgaben mit zunehmen-

der Übung zu einer schnelleren und akkurateren Leistung bei typgleichen Aufgaben mit konsistenten Anforderungen führt (Ackerman, 1987; Anderson, 1993). Wenn dieses repetitive Üben jedoch zu Lasten kognitiv herausfordernder Elemente des Mathematikunterrichts geht, wirkt es sich negativ auf die Gesamtleistung in mathematischen Schülerleistungstests aus (Gruehn, 2000).

Jedoch ist es schwierig, den Mathematikunterricht kognitiv herausfordernd zu gestalten, um mathematische Problemlösefähigkeiten und Verständnis zu fördern. Hiebert und Carpenter (1992, S. 65) sprechen hierbei zum Beispiel von „achieving this goal has been like searching for the Holy Grail“. Dennoch konnten einige Fortschritte in der Unterrichts- und Lehr-Lern-Forschung gemacht werden, die diese Lernziele erreichbarer machen (Baumert & Köller, 2000; Baumert u.a., 2004; De Corte, Greer & Verschaffel, 1996; Dixon, Carnine, Lee, Wallin & Chard, 1998; Gruehn, 2000; Klieme, Schümer u.a., 2001; Kunter, 2005; Schoenfeld, 1992). Ob Mathematikaufgaben verständnisvolle Lernprozesse oder das Lernen von Problemlösefähigkeiten stimulieren, ist im Rahmen der direkten Instruktion („Frontalunterricht“) vornehmlich von zwei Faktoren abhängig: den kognitiven Aufgabenanforderungen und wie kognitiv herausfordernd die Aufgaben im Unterricht behandelt werden bzw. wie die Lernaufgaben im Mathematikunterricht gestaltet sind (Klieme, Schümer u.a., 2001).

Die kognitiven Aufgabenanforderungen sollte die Lehrkraft im Vorfeld der Unterrichtsgestaltung mit dem Ziel analysieren (Mayer, 2004), ein detailliertes Verständnis für die lösungsrelevanten kognitiven Prozesse und das zur Lösung notwendige mathematische Wissen zu gewinnen, wie sie zum Beispiel im Modell mathematischen Problemlösens von Mayer (1992; vgl. Abschnitt 2.2) beschrieben werden (siehe auch Jordan u.a., 2006, für eine mathematikdidaktische Anforderungsanalyse von Aufgaben; Neubrand u.a., 2001).

Damit das kognitive Potenzial einer bestimmten Aufgabe im Rahmen der direkten Instruktion optimal zum Tragen kommt und die kognitive Eigenaktivität der Schüler gefordert und gefördert wird, stehen mehrere instruktionale Maßnahmen zur Verfügung (Baumert u.a., 2004):

- Beim Lösen von Mathematikaufgaben (multiples Lösungspotenzial vorausgesetzt) erlauben diese instruktionale Maßnahmen eine „Selbstdifferenzierung“ des kognitiven Anspruchs (vgl. Kraus u.a., 2004), da Schüler ihre Lösungswege selbst bestimmen können.
- Auch im Unterrichtsgespräch wird eine Selbstdifferenzierung angestrebt, da verschiedene (auch falsche Lösungsvarianten) von Schülern diskutiert werden, wobei jede Schülermeinung zählt und als wertvoller Beitrag in einer *community of practice* angesehen wird (Carpenter & Lehrer, 1999). Schüler sind bei diesen Diskussionen angehalten, ihre Lösungswege und ihr Vorgehen zu erklären, was eine tiefere Reflexion der zu Grunde liegenden mathematischen Konzepte, Operatoren und Lösungsstrategien stimuliert (vgl. Schoenfeld, 1985).
- Ein dritter Aspekt kognitiv-herausfordernder Unterrichtsarrangements ist das, was Gruehn (2000) als *anspruchsvolles Üben* bezeichnet. Dabei werden beim Üben die Anwendungskontexte mathematischer Operatoren, Schemata oder Lösungsstrategien variiert, was die Schüler anhält, ihr bisher gelerntes Wissen auf neue Kontexte zu transferieren (vgl. Carpenter & Lehrer, 1999).

Die Basisdimensionen der Unterrichtsqualität, repetitive, aber vor allem auch anspruchsvolle Formen des Übens und kognitiv-herausfordernde Unterrichtselemente, tragen zum Erreichen der Lehrplanziele bei. Ein entscheidender Aspekt bei der Analyse von Unterricht ist, dass es sich bei

diesen Unterrichtsmerkmalen um differenzielle Beschreibungsdimensionen handelt: Mathematikunterricht kann sich zwischen verschiedenen Schulklassen unterscheiden. Diese Unterschiede können auf unterschiedlichen Aggregationsniveaus, zum Beispiel Schulklassen, Einzelschulen oder Schulformen, zusammengefasst werden.

#### 5.4 Schulformen als differenzielle Entwicklungsmilieus

Eines der zentralen Merkmale des deutschen Schulsystems ist der gegliederte Aufbau mit dem Ziel einer Homogenisierung der Schülerschaft hinsichtlich ihrer Schülerleistung (Baumert, Trautwein & Artelt, 2003). Im deutschen Schulsystem gibt es verschiedene Formen der Leistungsdifferenzierung: eine Differenzierung in verschiedene Schulformen (Hauptschule, Realschule oder Gymnasium), bei Schulen mit mehreren Bildungsgängen sowie kooperativen Gesamtschulen eine schulinterne Leistungsdifferenzierung in verschiedene Bildungsgänge und bei integrierten Gesamtschulen eine fachspezifische Differenzierung nach Kursniveaus (vgl. Köller, 2003).

Mehrere Befunde aus der Schul-, Unterrichts- und Lehrerforschung sind mit Blick auf das gegliederte Schulsystem bedeutsam:

- Eine Differenzierung der Schülerschaft nach ihrer Leistung geht einher mit einer Differenzierung nach sozioökonomischen, psychosozialen, motivationalen und kognitiven Variablen (vgl. Baumert u.a., 2003).
- Einige Studien deuten darauf hin, dass Lehrer ihren Unterricht an das Leistungsniveau, wie auch die Leistungsheterogenität ihrer Klassen anpassen (Baumert u.a., 2004; Baumert, Schmitz, Sang & Roeder, 1987; Raudenbush, Rowan & Cheong, 1993; Weinert & Helmke, 1995b).
- Schulformen erklären den größten Varianzanteil der Mathematikleistung, der zwischen Schulen besteht. Dies konnten beispielsweise Baumert u.a. (2003) anhand von zwei Datensätzen für die Jahrgangsstufe 9 aus der PISA-2000-Studie und für die Jahrgangsstufen 7 und 10 aus der Studie *Bildungsverläufe und psychosoziale Entwicklung im Jugend- und jungen Erwachsenenalter* (BIJU) zeigen. Wie zu erwarten erzielen Schüler an der Hauptschule im Durchschnitt geringere Leistungen in Mathematik als Schüler an der Realschule. Auch liegt die mittlere mathematische Schülerleistung von Realschülern unter der mittleren Leistung von Gymnasiasten (siehe Tab. 15; Blum u.a., 2004; Hosenfeld, Köller & Baumert, 1999; Klieme, Neubrand u.a., 2001; Köller & Baumert, 2002; Kunter, 2005).
- Entwicklungsverläufe mathematischer Schülerleistung können sich zwischen den Schulformen unterscheiden: An Gymnasien sind (auch bei Kontrolle der Ausgangsleistungen) die stärksten Zugewinne festzustellen, gefolgt von Realschulen und Hauptschulen (Köller & Baumert, 2001, 2002).
- Lehrer in Deutschland unterscheiden sich in ihrer fachlichen Ausbildung für die unterschiedlichen Schulformen (Terhart, 2003). So nehmen auch Köller und Baumert (2002) an, dass die schulformspezifischen Leistungsentwicklungen in Mathematik Folge einer schulformspezifischen Lehrerausbildung sein könnten.

Nimmt man alle diese Befunde zusammen, ist es nahe liegend die Wirkung des Schulkontextes als Gegenstand von Forschungsfragen zu machen und Schulformen potenziell als differenzielle Entwicklungsmilieus mathematischer Schülerleistung zu betrachten (Baumert, Köller & Schnabel, 2000; Baumert u.a., 2003; Köller, Baumert & Schnabel, 2003). Eine Hauptdeterminante der Entwicklung mathematischer Schülerleistung ist der Mathematikunterricht. Daher ist es vor dem Hintergrund differenzieller Entwicklungsmilieus interessant zu fragen, wie sich der Mathematikunterricht zwischen verschiedenen Schulformen hinsichtlich leistungsförderlicher Aspekte unterscheidet. Nachfolgend werden einige aktuelle Befunde zu Unterschieden zwischen Hauptschulen, Realschulen und Gymnasien des Mathematikunterrichts aus bundesländerübergreifenden Studien für die Klassenstufen 7, 8 und 9 dargestellt. Diese drei Schulformen werden deshalb herausgegriffen, da hierbei die institutionalisierte Trennung der Schülerschaft mit der Leistungsdifferenzierung einhergeht.

Gruehn (2000) untersuchte die Unterrichtswahrnehmung von Schülern der 7. Jahrgangsstufe auf der Basis von Daten aus der BIJU-Studie (Baumert u.a., 2003). Beim Vergleich von Schülern an Gymnasien mit Schülern anderer Schulformen fand sie, dass an Gymnasien die Klassenführung effizienter ist, deutlich weniger repetitiv geübt wird und verstärkt kognitiv-aktivierende Unterrichtsformen (z.B. genetisch-sokratisches Vorgehen im Unterrichtsgespräch oder anspruchsvolles Üben) eingesetzt werden. Auch wurden an Gymnasien deutlich seltener binnendifferenzierende Maßnahmen oder Möglichkeiten der Schülermitbestimmung berichtet.

Klieme, Schümer und Knoll (2001) analysierten Videobeobachtungen des Mathematikunterrichts in der 8. Klasse, die im Rahmen der TIMS-Studie (Baumert, Bos & Lehmann, 2000) aufgezeichnet wurden. Ergebnis war, dass kognitiv-aktivierende Unterrichtselemente am häufigsten an Gymnasien und etwas weniger häufig an Realschulen beobachtet werden konnten. An Hauptschulen wurden diese Aspekte des Unterrichts am seltensten beobachtet. Unterschiede zwischen den Schulformen in der Effizienz der Klassenführung waren marginal. An Hauptschulen wurde eine stärkere Schülerorientierung der Lehrkräfte beobachtet.

Kunter (2005) reanalyisierte die Videoaufzeichnungen aus der TIMS-Studie und zog zusätzlich Schülerurteile des Mathematikunterrichts aus der 8. Klasse hinzu. Die von ihr untersuchten Unterrichtsmerkmale waren aktive Konstruktion (dies betrifft im Wesentlichen kognitiv-herausfordernde Elemente des Unterrichts), Relevanz von Mathematik im Alltag und Unterstützung der Eigenständigkeit der Lernprozesse (womit im weiteren Sinne explorative und wenig lehrergeleitete Aspekte des Unterrichts beschrieben werden), die sie mit Videobeobachtungen erfasste. Die Schülerurteile nutzte Kunter zur Beschreibung der Effektivität der Klassenführung, Mitbestimmungsmöglichkeiten und der Individualisierung im Unterricht. Zentrales Ergebnis der Studie war, dass sich die Schulformen deutlich auf den untersuchten Merkmalsdimensionen unterschieden. An Hauptschulen fand relativ selten aktive Konstruktion statt, Relevanz und Eigenständigkeit wurden betont, und es wurden verstärkt Möglichkeiten der Mitbestimmung und binnendifferenzierende Maßnahmen berichtet. Der Mathematikunterricht an den Gymnasien zeichnete sich durch vermehrte Möglichkeiten zur aktiven Konstruktion und mehr Störungsfreiheit aus, wohingegen weniger Gelegenheiten für eigenständiges Lernen eingeräumt wurden. Gymnasiasten berichteten auch weniger von Partizipationsmöglichkeiten und Elementen der Binnendifferenzierung. Die Realschule nahm eine Mittelposition zwischen den Polen Hauptschule und Gymnasien ein.



Im Rahmen der PISA-2000-Studie wurden Schüler aus mehreren Jahrgängen (wobei die meisten Schüler die 9. Klasse besuchten) nach ihrer Wahrnehmung des Mathematikunterrichts befragt (Klieme & Rakoczy, 2003). Schüler am Gymnasium berichteten von einer stärkeren kognitiven Aktivierung im Mathematikunterricht, von weniger Disziplinproblemen, aber auch gleichzeitig von geringerer Unterstützung seitens der Lehrkräfte. Hingegen berichteten Schüler an der Hauptschule von einer geringer ausgeprägten kognitiven Aktivierung im Mathematikunterricht, mehr Disziplinproblemen und mehr Unterstützung durch die Lehrkräfte. Schüler an der Realschule lagen wiederum zwischen den Polen Hauptschule und Gymnasium.

Baumert und Kollegen (2004) konnten auf Grundlage der Daten aus dem Projekt COACTIV (Krauss u.a., 2004) und Daten aus PISA 2003 auf die Urteile von Mathematiklehrkräften und die Schülerwahrnehmung des Unterrichts zurückgreifen. Lehrkräfte an Gymnasien berichteten im Vergleich zu den Hauptschullehrkräften von einer kognitiv-herausfordernden Unterrichtsgestaltung, einem weniger eng geführtem Unterricht<sup>12</sup>, weniger Individualisierungsmaßnahmen und von einer relativ effizienten Klassenführung. Die Urteile der Lehrkräfte an Realschulen lagen bezüglich der berichteten kognitiv-aktivierenden Unterrichtsgestaltung und Engführung des Unterrichts zwischen denen der Lehrkräfte an Hauptschulen und Gymnasien und bei der effektiven Klassenführung und Individualisierungsmaßnahmen gleichauf mit den Gymnasiallehrkräften. Dieses Befundmuster zeigte sich auch bei Kontrolle der mittleren Mathematikleistung der jeweiligen Klassen, was nach Baumert und Kollegen für relativ stabile institutionelle Unterrichtskulturen spricht. Ein weiterer interessanter Befund in der Studie von Baumert und Kollegen war, dass Hauptschüler ihren Unterricht als deutlich stärker kognitiv-herausfordernd und selbstständigkeitsfördernd wahrnahmen als Realschüler und Schüler an Gymnasien. Hingegen beurteilten die Hauptschüler die Klassenführung als deutlich weniger effizient als an den beiden anderen Schulformen. An Realschulen und Gymnasien beurteilte die Schülerschaft diese beiden Unterrichtsaspekte nahezu gleich.

Zusammenfassend soll die Eingangsfrage „Wie unterscheidet sich das Angebot ‚Mathematikunterricht‘ zwischen verschiedenen Schulformen?“ beantwortet werden. Es ergibt sich ein relativ klares Bild zu schulformspezifischen Unterschieden.<sup>13</sup> Dieses kann pointiert wie folgt beschrieben werden: An Gymnasien wird kognitiv-herausfordernd, wenig individualisierend und schülerorientiert und störungspräventiv unterrichtet. An Hauptschulen hingegen zeichnet sich der Mathematikunterricht durch geringere kognitive Herausforderung, eine erhöhte Anfälligkeit gegenüber Störungen und einer verstärkten Individualisierung und Schülerorientierung aus. Der Mathematikunterricht an Realschulen liegt zwischen diesen beiden Polen.

Der referierte Hintergrund zu schulformspezifischen Unterschieden des Mathematikunterrichts unterfüttert die zweite Forschungsfrage: „Wie unterscheiden sich Schüler, die unterschiedliche Schulformen besuchen, hinsichtlich der kognitiven Fähigkeiten mathematischer Schülerleistung?“ Auf Grundlage der leistungsförderlichen Unterschiede des Mathematikunterrichts lassen sich einerseits Hypothesen ableiten, wie sich Schüler im Mittel in den kognitiven Fähigkeiten unterscheiden. Andererseits können angesichts der theoretischen Überlegungen, dass Unterschiede

---

<sup>12</sup> Baumert und Kollegen verstehen unter eng geführtem Unterricht unter anderem Unterrichtselemente, die als einschleifendes repetitives Üben, verständnissicherndes remediales Wiederholen, als eine kleinschrittige Dokumentation der Lösungswege oder eine fürsorgliche Anleitung und Kontrolle beschrieben werden können.

<sup>13</sup> Allerdings ist es nicht gerechtfertigt, allein die Schulformunterschiede des Mathematikunterrichts als Ursache für schulformspezifische Effekte des Leistungszuwachses zu betrachten. Wie schon die Modelle schulischen Lernens nahe legen, ist Lernen vom komplexen Zusammenspiel kognitiver und motivationaler Schülervariablen und der Unterrichtsqualität abhängig.

zwischen den Lernumwelten „Mathematikunterricht“ (im Sinne von unterschiedlichen Investitionsmöglichkeiten für die allgemeine kognitive Fähigkeit) zur Differenzierung mathematischer Fähigkeiten führen können (Abschnitt 4.3.1 und 4.3.2), auch Hypothesen zu schulformspezifischen Unterschieden der Differenzierung spezifiziert werden. Eine genaue Formulierung dieser Hypothesen findet sich in Kapitel 8.

## 6 Zusammenfassung vor dem Hintergrund der Forschungsfragen

Im Zentrum dieser Arbeit steht ein Forschungsansatz im Sinne des *coordinate measurement* (Messick, 1984) von Maßen kognitiver Fähigkeiten und Maßen mathematischer Schülerleistung. Im Rahmen dieses integrativen Ansatzes werden drei Forschungsfragen untersucht, vor deren Hintergrund nachfolgend die Theorien und Befunde der vorangegangenen Kapitel nochmals knapp zusammengefasst werden. Konkrete Hypothesen werden in den Kapiteln zu den Forschungsfragen des empirischen Teils abgeleitet und soweit es möglich ist durch die Referenz empirischer Befunde gestützt.

*Forschungsfrage 1: Struktur mathematischer Schülerleistung.* Eine Interpretation des *coordinate measurement* ist, Maße mathematischer Schülerleistung und kognitive Fähigkeiten in einem gemeinsamen Strukturmodell zu analysieren. Mögliche Strukturmodelle wurden im Rahmen von Abschnitt 4.2 erörtert. Ein zentrales Merkmal dieser Modelle ist der hierarchische Aufbau der Fähigkeitskonstrukte, von dem alle moderneren Strukturmodelle ausgehen. Dies impliziert, dass eine globalere mathematische Fähigkeit spezifischeren mathematischen Fähigkeiten hierarchisch übergeordnet ist.

Aber einige Fragen zur Struktur sind offen, wie die diskutierten Unterschiede zwischen den Modellen zeigten (siehe auch Tab. 2). Im Einzelnen sind das die Annahme einer allgemeinen kognitiven Fähigkeit (Gf-Gc-Theorie vs. Vernon, Carroll, BIS-Modell), die Annahme multipler Bedingtheit der Aufgaben (Gf-Gc-Theorie vs. Vernon, Carroll, BIS-Modell) und die Konzeption mathematischer Fähigkeit (operativ vs. inhaltlich).

Eine genaue Explikation dieser drei Unterschiede erfolgt in den Abschnitt 7.1 und 7.2. Dabei wird die zugehörige empirische Befundlage dargestellt, und es werden Hypothesen abgeleitet, die in Form von Strukturmodellen geprüft werden.

Die Ergebnisse dieser Modellprüfungen fokussieren auf das erste Hauptziel der Arbeit: eine Anbindung mathematischer Schülerleistung an differenzialpsychologische kognitive Fähigkeitskonstrukte (vgl. Köller & Baumert, 2002). Mit dieser Anbindung werden die Quellen interindividueller Unterschiede mathematischer Schülerleistung in Termini psychologischer Fähigkeitskonstrukte spezifiziert. Die nächsten beiden Forschungsfragen setzen an diesem Punkt an und untersuchen Schulformunterschiede und die externe Validität der kognitiven Fähigkeiten mathematischer Schülerleistung.

*Forschungsfrage 2: Schulformunterschiede.* Die zweite Forschungsfrage kann in zwei Teilaspekte untergliedert werden. Der *erste Teilaspekt* konzentriert sich auf Schulformunterschiede hinsichtlich des Niveaus der kognitiven Fähigkeiten mathematischer Schülerleistung. Zur Beantwortung der Frage nach Niveauunterschieden sind vor allem zwei Überlegungen relevant:

- Der gegliederte Aufbau des deutschen Schulsystems strebt eine Leistungsdifferenzierung der Schülerschaft zwischen den Schulformen an. Personen mit höherem Niveau kognitiver Fähigkeiten besuchen in der Regel auch höhere Schulformen. So sprechen beispielsweise Baumert, Trautwein und Artelt (2003, S. 283) von allgemeiner kognitiver Fähigkeit (sie bezeichnen diese als kognitive Grundfähigkeiten) als zentrales Merkmal der „Übergangsauslese“: Das Niveau allgemeiner kognitiver Fähigkeit bestimmt somit den Übergang in höhere Schulformen maßgeblich mit.

- In höheren Schulformen sind tendenziell lernförderliche Aspekte des Mathematikunterrichts (z.B. „Störungsprävention“ und „kognitive Herausforderung“) auf höherem Niveau ausgeprägt (Abschnitt 5.4).

Nimmt man diese beiden Befunde zusammen, dann sind für Schüler unterschiedlicher Schulformen Niveauunterschiede der kognitiven Fähigkeiten zu erwarten, die mathematischer Schülerleistung zu Grunde liegen. Die Frage nach Niveauunterschieden zwischen den Schulformen ist natürlich nicht neu (Blum u.a., 2004; Klieme, Neubrand & Lüdtke, 2001; Köller & Baumert, 2002; Kunter, 2005). Dennoch ist sie interessant, weil in bisherigen Studien zu schulformspezifischen Niveauunterschieden zwei zentrale Aspekte nicht beachtet wurden:

- erstens die Berücksichtigung rivalisierender Strukturmodelle der kognitiven Fähigkeiten mathematischer Schülerleistung (vgl. Abschnitt 4.2 und Forschungsfrage 1),
- zweitens die Prüfung der zentralen Voraussetzung, ob die verwendeten Modelle überhaupt geeignet sind, schulformübergreifend die Struktur mathematischer Schülerleistung zu beschreiben.

Die Teilstudie zu schulformspezifischen Niveauunterschieden wird versuchen, diese beiden Forschungslücken zu schließen.

Der *zweite Teilaspekt* der zweiten Forschungsfrage fokussiert auf die Heterogenität der kognitiven Fähigkeiten mathematischer Schülerleistung. Hierzu sind drei Befunde und Überlegungen zur Differenzierung kognitiver Fähigkeiten relevant:

- Unterschiede in der Lernumwelt „Mathematikunterricht“ können zur Differenzierung kognitiver Fähigkeiten führen (vgl. Abschnitt 4.3.1). Hierbei sind insbesondere schulformspezifische Unterschiede des Mathematikunterrichts relevant (vgl. Abschnitt 5.4).
- Personen mit höherem Niveau allgemeiner kognitiver Fähigkeit weisen eine differenziertere Struktur kognitiver Fähigkeiten auf als Personen mit niedrigerem Fähigkeitsniveau (Abschnitt 4.3.2). So legen die Ergebnisse der meisten referierten Studien nahe, dass mathematische Fähigkeiten mit zunehmender allgemeiner kognitiver Fähigkeit differenzieren.
- Abad, Colom, Juan-Espinosa und Garcia (2003) erklären die Differenzierung mit zunehmendem Niveau allgemeiner kognitiver Fähigkeit (unter anderem) dadurch, dass Personen mit höherem Fähigkeitsniveau Zugang zu besseren Bildungsmöglichkeiten (z.B. höheren Schulformen) haben und somit sensu Jensen (1998) bessere Investitionsmöglichkeiten für ihre allgemeine kognitive Fähigkeit vorfinden.

Vor dem Hintergrund dieser Befunde und Überlegungen ist es interessant zu fragen, ob die Differenzierung der kognitiven Fähigkeiten mathematischer Schülerleistung von der Schulform abhängig ist. Diese Hypothese wurde in früheren Studien zur Differenzierung kognitiver Fähigkeiten bisher noch nicht geprüft.

Zusammenfassend kann festgehalten werden, dass die Analysen zu schulformspezifischen Unterschieden im Niveau und in der Heterogenität der kognitiven Fähigkeiten mathematischer Schülerleistung dabei helfen, einen sehr detaillierten und methodisch fundierten Einblick in Schulformunterschiede zu erhalten.

*Forschungsfrage 3: Validität.* Zusätzlich zu Schulformunterschieden ist es wichtig, den Zusammenhang zwischen den kognitiven Fähigkeiten mathematischer Schülerleistung und mehreren Außenkriterien zu untersuchen, um diese Fähigkeiten inhaltlich weiter zu unterfüttern. Die entsprechenden Analysen werden im Rahmen der dritten Forschungsfrage durchgeführt. Die vorliegende Arbeit ist natürlich nicht die erste, die Fragen zur (externen) Validität mathematischer Schülerleistung untersucht. Zum Beispiel ist es ein vielfach replizierter Befund, dass mathematische Schülerleistung und mathematikspezifische Lernmotivation (z.B. mathematisches Selbstkonzept oder mathematisches Interesse) positiv kovariieren (Abschnitt 5.2.2).

Allerdings muss man bei diesen Befunden beachten, dass bei nahezu allen Analysen alternative Konzeptualisierungen kognitiver Fähigkeiten mathematischer Schülerleistung *nicht* in Betracht gezogen wurden (für einige Ausnahmen siehe Gustafsson & Snow, 1997). Dies erstaunt, da – wie bei der ersten Forschungsfrage erläutert wurde – einige Fragen zur strukturellen Repräsentation dieser Fähigkeiten offen sind.

In den bisherigen Analysen wurde ein *Standardmodell* verwendet, das mehr oder weniger explizit auf den Annahmen des Gf-Gc-Modells (Horn & Noll, 1997) aufbaute. In den Validitätsanalysen wurde dann in der Regel ein (manifestes, messfehlerbehafteter) globaler Testscore mathematischer Schülerleistung (Gustafsson & Snow, 1997, S. 127) als Maß für eine generelle mathematische Fähigkeit (M, Abb. 13a) mit verschiedenen (meist ebenfalls messfehlerbehafteten) Schülermerkmalen interkorreliert (Reeve, 2004, S. 624–625).

Dieses methodische Vorgehen kann aber problematisch sein, wie Reeve deutlich macht:

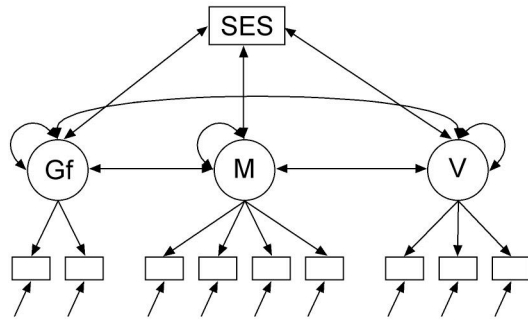
By ignoring the fact that most test subscales confound sources of variance due to multiple specific and general abilities, many prior studies have failed to obtain reliable and construct-valid assessments of narrow abilities. Indeed, studies of the validities of narrow abilities often estimate these factors with the variance due to g included. Thus, the correlations with outcomes reflect both the variance due to g as well as the unique variance due to the specific factor (...) The failure to parse the variance in observed tests into meaningful sources, based on a validated model of the structure of abilities, raises questions about the validity of the operationalization of narrow abilities in some of the past research. (Reeve, 2004, S. 625)

Was bedeutet das für den zu erwartenden Zusammenhang von mathematischer Schülerleistung und Schülermerkmalen? Folgt man dem Argument von Reeve (siehe auch Gustafsson & Balke, 1993; Gustafsson & Snow, 1997), dann können sich die Korrelationen mathematischer Schülerleistung mit Außenkriterien in Abhängigkeit des gewählten Strukturmodells mathematischer Schülerleistung unterscheiden. Hierbei ist insbesondere der Vergleich für zwei unterschiedliche Konzeptualisierungen der kognitiven Fähigkeiten mathematischer Schülerleistung interessant.

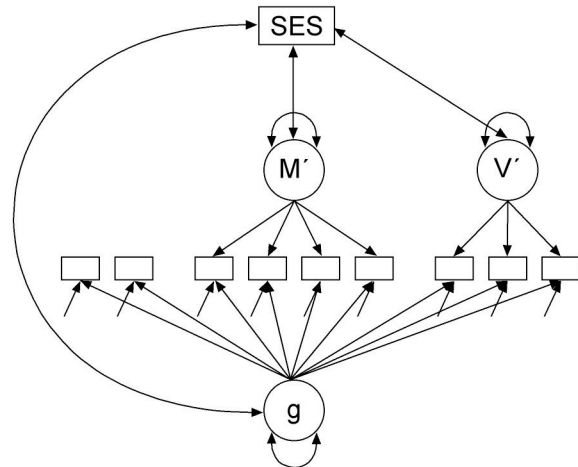
Das erste Modell ist das *Standardmodell*, in dem eine generelle mathematische Fähigkeit (M) die interindividuellen Unterschiede in den Maßen mathematischer Schülerleistung erklärt. Im Gegensatz hierzu erklären im zweiten Modell (Abb. 13b), einem so genannten *Nested-Faktormodell* (Gustafsson & Balke, 1993), die allgemeine kognitive Fähigkeit (g) und eine mathematikspezifische Fähigkeit (M') interindividuelle Unterschiede bei Maßen mathematischer Schülerleistung (vgl. die Annahmen von Vernon, von Carrolls Drei-Stratum-Theorie, und des BIS-Modells von Jäger u.a.). Damit ist es möglich, die korrelativen Beziehungen der allgemeinen kognitiven Fähigkeit und der mathematikspezifischen Fähigkeit (z.B. zu Maßen des familiären Hintergrunds: „SES“ in Abb. 13) getrennt voneinander zu untersuchen. Dies ist insbesondere dann von Interesse, wenn die allgemeine kognitive Fähigkeit andere korrelative Beziehungen zu den

Außenkriterien aufweist als die mathematikspezifische Fähigkeit (konkrete Hypothesen hierzu werden in Kap. 9 formuliert). In diesen Fällen ist davon auszugehen, dass sich die Korrelationen mit Außenkriterien zwischen  $M$  und  $M'$  unterscheiden. Der Grund hierfür ist, dass in der Varianz der generellen mathematischen Fähigkeit Varianzanteile der allgemeinen kognitiven Fähigkeit und der mathematikspezifischen Fähigkeit gemischt (*confounded* sensu Reeve, 2004) enthalten sind.

(a) Standard Gf-Gc-Modell



(b) Nested-Faktormodell



SES = sozioökonomischer Status, Gf = fluide Fähigkeiten, M = generelle mathematische Fähigkeiten, V = verbale Fähigkeit,  $M'$  = mathematikspezifische Fähigkeit,  $V'$  = spezifische verbale Fähigkeit, g = allgemeine kognitive Fähigkeit.

Abbildung 13: Analyse der differentiellen Validität der kognitiven Fähigkeiten mathematischer Schülerleistung

Die Analysen im Rahmen der dritten Forschungsfrage erlauben es, zwei Ziele zu verfolgen:

- Durch die Verwendung des Standardmodells wird versucht, die Befunde zur externen Validität mathematischer Schülerleistung aus bisherigen Studien zu replizieren.
- Bei Konzeptualisierung der kognitiven Fähigkeiten mathematischer Schülerleistung in Form des Nested-Faktormodells kann (möglicherweise) ein – im Vergleich zum bisherigen Vorgehen – genaueres Bild der Zusammenhänge zwischen den kognitiven Fähigkeiten mathematischer Schülerleistung und soziodemografischen und motivationalen Schülermerkmalen sowie Schulnoten gezeichnet werden.

In den nachfolgenden Kapiteln werden nun drei Teilstudien vorgestellt, die empirische Antworten auf die drei Forschungsfragen der vorliegenden Arbeit geben.

## 7 Studie 1: Struktur mathematischer Schülerleistung

Das Hauptziel der vorliegenden Arbeit ist, mathematische Schülerleistung als psychologisches Konstrukt zu spezifizieren, um sie aus dieser Perspektive besser zu verstehen. In dieser Arbeit wird dieses Ziel aus einer differenzialpsychologischen Sicht verfolgt: Die erste Forschungsfrage beschäftigt sich mit der Frage, welche kognitiven Fähigkeiten interindividuelle Unterschiede bei Aufgaben zur Messung mathematischer Schülerleistung erklären können. Die Antwort erfolgt in zwei Schritten. Diesem Vorgehen liegt die Annahme zu Grunde, dass kognitive Fähigkeitskonstrukte hierarchisch organisiert sind und eine globalere mathematische Fähigkeit spezifischeren mathematischen Fähigkeiten übergeordnet ist. In Abschnitt 7.1 werden zunächst diese spezifischen mathematischen Fähigkeiten untersucht. Abschnitt 7.2 befasst sich mit der nächsten Hierarchieebene und beleuchtet das Zusammenspiel von mathematischen Fähigkeiten und kognitiven (nicht genuin mathematischen) Fähigkeiten (vgl. Abschnitt 4.2).

Beide Unterkapitel sind in gleicher Weise aufgebaut. Zunächst wird die empirische Befundlage dargestellt. Auf dieser Grundlage und anhand der referierten Strukturmodelle (vgl. Abschnitt 4.2) werden Hypothesen abgeleitet, welche mathematischen Fähigkeiten unterschieden werden können bzw. welche weiteren (nicht genuin mathematischen) Fähigkeiten interindividuelle Unterschiede bei Aufgaben zur Messung mathematischer Schülerleistung erklären können. Die Hypothesen werden in Form von *Strukturmodellen* repräsentiert. Im anschließenden Methodenteil werden jeweils die verwendeten Schülerstichproben und Messinstrumente beschrieben und das analytische Vorgehen dargestellt. Im Anschluss werden die Ergebnisse der Modellprüfungen dokumentiert und kurz diskutiert. Abschließend folgt eine Gesamtdiskussion zur ersten Forschungsfrage.

### 7.1 Mathematische Fähigkeiten

#### 7.1.1 Empirische Befundlage und zu prüfende Strukturmodelle

Welche mathematischen Fähigkeiten können die interindividuellen Unterschiede bei Aufgaben zur Messung mathematischer Schülerleistung erklären? Ein zentraler Befund bei der Diskussion der Strukturmodelle kognitiver Fähigkeiten ergab, dass im Gf-Gc-Modell (Horn & Noll, 1997) und im BIS-Modell (Jäger, Süß & Beauducel, 1997) eine operative Unterscheidung mathematischer Fähigkeiten vorgenommen wird. Vernon (1964) differenziert in seinem Modell hingegen die mathematische Fähigkeit inhaltlich nach mathematischen Stoffgebieten. Carroll (1963) nimmt im Rahmen der Drei-Stratum-Theorie eine integrative Position ein und unterscheidet sowohl operativ (Rechenfertigkeit vs. quantitatives Reasoning) als auch innerhalb von quantitativem Reasoning nochmals inhaltlich nach Stoffgebieten. Gemein ist allen vier Modellen, dass sie von einem hierarchischen Aufbau ausgehen. Operative oder stoffgebietsspezifische mathematische Teilfähigkeiten sind einer globaleren mathematischen Fähigkeit untergeordnet. Wie sieht die empirische Befundlage hierzu aus? Nachfolgend werden einige ausgewählte Ergebnisse aus *Large-Scale*-Studien vorgestellt.

*Empirische Befundlage.* Gustafsson (1994) analysierte die Struktur mathematischer Schülerleistung bei einer Stichprobe schwedischer Fünftklässler. Dabei differenzierte er mathematische Schülerleistung operativ. Er unterschied mathematisches Problemlösen (Aufgaben erforderten das Lösen von Textaufgaben), numerisches Verständnis (Aufgaben erforderten das verständnisvolle Anwenden numerischer Begriffe) und arithmetische Gleichungen (Aufgaben erforderten das Lösen

arithmetischer Gleichungen). Die drei Faktoren, die die operativen Facetten repräsentierten, interkorrelierten dabei hoch miteinander. Die Korrelationen lagen zwischen .74 und .90. Ein hierzu alternatives Nested-Faktormodell sah eine generelle Mathematikfähigkeit sowie davon unabhängige spezifische Fähigkeiten für numerisches Verständnis und arithmetische Gleichungen vor. Beide Modelle wiesen guten Modell-Fit auf, wobei das Nested-Faktormodell einen etwas besseren Fit hatte.

In der US-amerikanischen NELS: 88-Längsschnittstudie bearbeiteten Schüler in der 8. Klasse, in der 10. Klasse und in der 12. Klasse mathematische Schülerleistungstests. Bei der Zusammenstellung der Tests wurden sowohl mathematische Stoffgebiete als auch kognitive Operationen berücksichtigt. Im Rahmen von zwei Arbeiten reanalysierten Kupermintz und Kollegen (Kupermintz, Ennis, Hamilton, Talbert & Snow, 1995; Kupermintz & Snow, 1997) mit exploratorischen Itemfaktorenanalysen (Bock, Gibbons & Muraki, 1988) die Daten von NELS: 88. In allen drei Jahrgangsstufen konnten sie einen mathematischen Wissensfaktor extrahieren. Dieser repräsentierte nach Kupermintz und Kollegen vornehmlich die Fähigkeit zur Ausführung (einfach zu lösender) mathematischer Berechnungen sowie die Fähigkeit zur Anwendung mathematischer Begriffe und Algorithmen. Ein zweiter replizierbarer Faktor mathematischen Schlussfolgerns repräsentierte die Fähigkeit, komplexe mathematische Probleme zu lösen. In der 8. Klasse korrelierten die beiden Faktoren mit  $r = .72$ , in der 10. Klasse mit  $r = .75$  und in der 12. Klasse mit  $r = .67$ .

Muthén, Khoo und Goff (1997) reanalysierten den Datensatz aus dem Jahr 1992 für die Jahrgangsstufen 8 und 12 aus der US-amerikanischen Studie *National Assessment of Educational Progress* (NAEP). Muthén und Kollegen spezifizierten dabei zwei Faktormodelle mathematischer Schülerleistung. Ein Modell sah interkorrelierte Faktoren für fünf mathematische Stoffgebiete (Zahlen und Operationen, Messung, Algebra, Geometrie und Stochastik) vor. Die latenten Interkorrelationen der stoffgebietsspezifischen Fähigkeiten lagen dabei für Schüler in der 8. Klasse zwischen 0,84 und 0,99 und für Schüler in der 12. Klasse zwischen 0,95 und 1,0. Im zweiten Modell wurde ein Nested-Faktormodell spezifiziert, das eine generelle mathematische Fähigkeit und hiervon unabhängige stoffgebietsspezifische Fähigkeiten vorsah. Von Interesse war dabei der Varianzanteil der jeweiligen Fähigkeiten an der reliablen Gesamtvarianz. Die generelle mathematische Fähigkeit erklärte in beiden Jahrgangsstufen etwa 80 Prozent der reliablen Varianz auf. Die Varianzanteile der stoffgebietsspezifischen Fähigkeiten lagen zwischen 0 und 25 Prozent.

Für die deutsche TIMSS-Schülerstichprobe der 7. Jahrgangsstufe berichtete Köller (1998) die Interkorrelationen zwischen sechs mathematischen Stoffgebieten. Die Interkorrelationen der raschskalierten Personenparameter lagen dabei zwischen .31 (Proportionalität mit Messen/Maßeinheiten) und .53 (Zahlen/Zahlenverständnis mit Algebra).<sup>14</sup> Weiterhin konnte Köller auf Grundlage der Personenparameter einen Faktor extrahieren, der eine generelle mathematische Fähigkeit repräsentierte.

Im Rahmen von TIMSS wurden auch deutsche Schüler in der gymnasialen Oberstufe untersucht. Die Mathematikaufgaben des TIMSS-Tests zur „voruniversitären Mathematik“ wurden dabei sowohl inhaltlich nach mathematischen Stoffgebieten als auch operativ nach kognitiven Anforder-

---

<sup>14</sup> Die im Vergleich zu den anderen Studien deutlich niedrigeren Interkorrelationen sind darauf zurückzuführen, dass Köller bei seinen Analysen Personenparameter verwendete. Diese stellen nur eine Schätzung der latenten Personenfähigkeit dar und führen somit aufgrund der nicht perfekten Reliabilität der Personenparameter (siehe z.B. Rost, 2004a) zu verminderten (attenuierten) Schätzungen der Interkorrelationen zwischen den mathematischen Stoffgebieten.



rungen klassifiziert (Klieme, 2000). Die latenten Interkorrelationen für die Stoffgebiete lagen zwischen .77 (Geometrie mit Analysis) und .81 (Zahlen, Gleichungen und Funktionen mit Analysis, Zahlen, Gleichungen und Funktionen mit Geometrie). Die latente Korrelation zwischen den operativen Fähigkeiten Anwenden/Problemlösen und Routineverfahren wie auch die Korrelation zwischen Anwenden/Problemlösen und komplexe Verfahren lag bei .82. Routineverfahren und komplexe Verfahren interkorrelierten mit .87.

Auf der Grundlage der Daten von 15-jährigen Schülern wurden bei PISA 2000 zwei Modelle mathematischer Fähigkeit geprüft, bei denen inhaltlich nach mathematischen Stoffgebieten und operativ nach den Arten mathematischen Arbeitens unterschieden wurde (Klieme, Neubrand u.a., 2001). In beiden Modellen lagen die berichteten latenten Interkorrelationen (mit Ausnahme der Interkorrelationen von Stochastik mit den anderen mathematischen Stoffgebieten) zwischen .88 und .96. Die Korrelationen von Stochastik mit Arithmetik, Algebra und Geometrie lagen zwischen .80 und .83.

Die gleichen Modelle wurden auch auf Grundlage der Daten 15-jähriger Schüler aus der PISA-2003-Stichprobe spezifiziert (Blum u.a., 2004). Hier lagen die latenten Interkorrelationen bei einer Unterscheidung nach mathematischen Stoffgebieten zwischen .82 (Stochastik und Geometrie) und .91 (Arithmetik und Algebra). Bei der operativen Unterscheidung nach den Arten mathematischen Arbeitens lagen die Interkorrelationen für technische Aufgaben und rechnerisches Modellieren sowie für technische Aufgaben und begriffliches Modellieren bei .89. Rechnerisches Modellieren korrelierte mit begrifflichem Modellieren mit .95.

*Zusammenfassung.* Zusammenfassend kann festgehalten werden, dass die referierten Befunde nationen- und jahrgangsstufenübergreifend relativ einheitlich sind: Mathematikspezifische Fähigkeiten waren hoch miteinander interkorreliert, unabhängig davon, ob diese auf einer operativen oder inhaltlichen Klassifikation beruhten (Blum u.a., 2004; Gustafsson, 1994; Klieme, 2000; Klieme, Neubrand u.a., 2001; Kupermintz u.a., 1995; Kupermintz & Snow, 1997; Muthén u.a., 1997). Weiterhin wurde in einigen Studien ein Faktor extrahiert (Köller, 1998) oder a priori spezifiziert (Gustafsson, 1994; Muthén u.a., 1997), der eine globalere mathematische Fähigkeit repräsentierte. Interkorrelationen, wie auch der empirisch replizierte Befund eines Faktors, der eine globalere mathematische Fähigkeit repräsentierte, sprechen eindeutig dafür, dass mathematische Teilfähigkeiten hierarchisch organisiert sind.

Im Gegensatz hierzu ist die Befundlage zur Unterscheidung operativer oder stoffgebietspezifischer mathematischer Fähigkeiten *nicht* eindeutig. Beide Unterscheidungen werden von den referierten Studien gestützt. Bemerkenswert ist, dass nur in drei der referierten Studien (Blum u.a., 2004; Klieme, 2000; Klieme, Neubrand u.a., 2001) Modelle für die inhaltliche *und* operative Differenzierung konfirmatorisch geprüft wurden. Leider wurde in keiner dieser Studien dokumentiert, welches der beiden Modelle den besseren Modell-Fit aufwies.

Zudem wurden im Rahmen von PISA bei beiden Erhebungszyklen die Stichproben 15-jähriger Schüler zur Modellprüfung herangezogen. Dies ist nicht unproblematisch, da diese Schüler unterschiedliche Klassenstufen besuchten. Wenn bestimmte mathematische Inhalte erst in höheren Klassenstufen unterrichtet werden, hatten einige Schüler noch gar keine Möglichkeit, diese Inhalte zu lernen. Damit sind interindividuelle Unterschiede in den mathematischen Fähigkeiten möglicherweise konfundiert mit dem Besuch unterschiedlicher Klassenstufen. Die Frage, ob die Ergebnisse aus PISA auf eine jahrgangshomogene Population generalisiert werden können, ist damit noch offen.

Zusammengenommen können auf Grundlage der referierten Befunde sowie auf Basis der theoretischen Konzeptionen in den Strukturmodellen mathematischer Fähigkeit (vgl. Abschnitt 4.2) drei (teilweise rivalisierende) Annahmen abgeleitet werden:

- Mathematische Schülerleistung ist hierarchisch organisiert.
- Mathematische Schülerleistung kann anhand von operativen mathematischen Fähigkeiten binnendifferenziert werden.
- Mathematische Schülerleistung kann anhand von stoffgebietsspezifischen mathematischen Fähigkeiten binnendifferenziert werden.

*Zu prüfende Strukturmodelle.* Zur empirischen Prüfung dieser Annahmen werden in der vorliegenden Arbeit drei Strukturmodelle mathematischer Schülerleistung a priori spezifiziert (Abb. 14):<sup>15</sup> Interindividuelle Unterschiede bei Aufgaben zur Messung mathematischer Schülerleistung werden darin durch interindividuelle Unterschiede bei einer (oder mehreren) latenten mathematischen Fähigkeit(en) erklärt.

*Modell 1: Generelle mathematische Fähigkeit.* Dieses Modell sieht nur eine generelle mathematische Fähigkeit vor. Dies korrespondiert mit der Annahme des hierarchischen Aufbaus mathematischer Fähigkeit. Spezifische mathematische Fähigkeiten werden durch die Residualterme der manifesten Mathematikaufgaben repräsentiert, da auch in den Residualvarianzen „wahre“ reliable Varianzanteile enthalten sein können (z.B. Bollen, 1989; Spearman, 1927). Da die Residualterme aufgabenspezifische Fähigkeiten repräsentieren, sind sie wechselseitig unkorreliert.

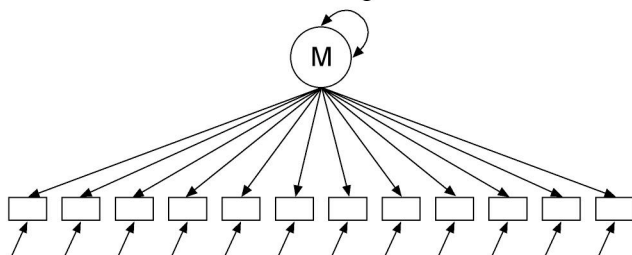
Mit diesem einfaktoriellen Modell kann zusätzlich die Frage beantwortet werden, ob bei Differenzierung operativer oder mathematischer Teilfähigkeiten die empirischen Daten besser erklärt werden können als mit Modell 1. Der Vergleich mit Modell 1 stellt somit auch einen empirischen Bezugspunkt für die Modelle 2 und 3 dar.

*Modell 2: Operative mathematische Teilfähigkeiten/Arten mathematischen Arbeitens.* In diesem Modell werden Mathematikaufgaben nach den Arten mathematischen Arbeitens klassifiziert. Somit werden drei operative Teilfähigkeiten mathematischer Schülerleistung unterschieden: technische Fertigkeiten (TF), die Fähigkeit zum rechnerischen Modellieren (RM) und die Fähigkeit zum begrifflichen Modellieren (BM). Die Annahme einer generellen mathematischen Fähigkeit wird durch die Interkorrelationen der drei operativen Teilfähigkeiten repräsentiert. Technische Fertigkeiten (in den PISA-Studien wurde hierfür der Begriff technische Aufgaben verwendet) werden durch Aufgaben gemessen, bei denen die mathematischen Ansätze vorgegeben sind und mittels bekannter Prozeduren bearbeitet werden können. Für rechnerische Modellierungsaufgaben müssen mathematische Lösungsansätze erstellt werden, die rechnerisch-prozedural aufzulösen sind. Aufgaben zur Messung von begrifflichem Modellieren sind dadurch charakterisiert, dass das Herstellen von begrifflichen Zusammenhängen zwischen mathematischen Konzepten lösungsrelevant ist (Blum u.a., 2004; Klieme, Neubrand u.a., 2001).

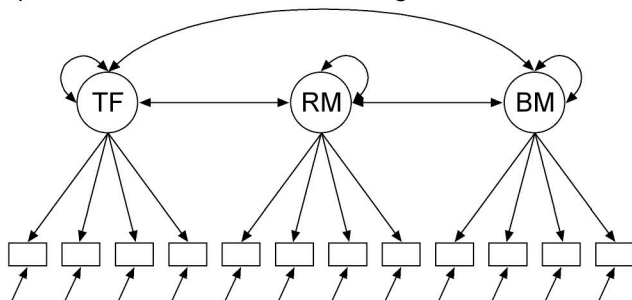
---

<sup>15</sup> In dieser Arbeit werden latente Variablen durch Kreise, manifeste Variablen durch Rechtecke und Residualterme durch Pfeile auf manifeste bzw. latente Variablen dargestellt. Korrelationen werden durch Doppelpfeile und Regressionen durch gerichtete Pfeile repräsentiert.

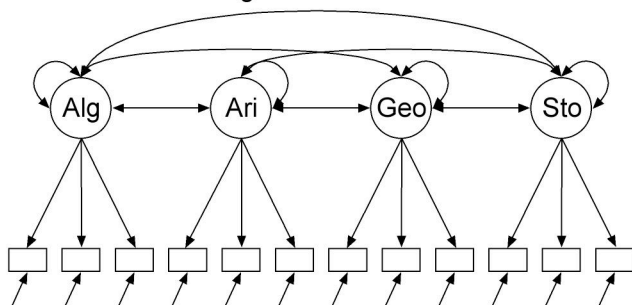
Modell 1: Generelle mathematische Fähigkeit



Modell 2: Operative mathematische Teilfähigkeiten/Arten mathematischen Arbeitens



Modell 3: Mathematische Stoffgebiete



M = generelle mathematische Fähigkeit, TF = technische Fertigkeiten, RM = rechnerisches Modellieren, BM = begriffliches Modellieren, Alg = Algebra, Ari = Arithmetik, Geo = Geometrie, Sto = Stochastik.

Abbildung 14: Mathematische Fähigkeiten

Warum wird diese operative Unterscheidung mathematischer Teilfähigkeiten gewählt? Die Umschreibung der Arten mathematischen Arbeitens korrespondiert eng mit der Konzeption mathematischer Fähigkeit in den Strukturmodellen (Gf-Gc-Modell, BIS-Modell, eingeschränkt auch die Drei-Stratum-Theorie von Carroll; vgl. Abschnitt 4.2), die mathematische Fähigkeit operativ differenzieren. So werden in allen Modellen problemorientierte und fertigungsorientierte Teilfähigkeiten unterschieden. Weiterhin wird sogar die Unterscheidung von genau diesen drei Arten mathematischen Arbeitens konzeptionell gestützt. So erfordern im Gf-Gc-Modell (Horn & Noll, 1997) Aufgaben zur Messung „mathematischen Wissens“ Verständnis und Anwendung mathematischer Konzepte (dies korrespondiert eng mit begrifflichem Modellieren) sowie die Anwendung mathematischer Fertigkeiten (dies korrespondiert eng mit technischen Fertigkeiten). Weiterhin wird mathematisches Wissen auch durch die Fähigkeit zum Lösen algebraischer Probleme gemessen (dies korrespondiert eng mit rechnerischem Modellieren).

*Modell 3: Mathematische Stoffgebiete.* Alternativ zur Unterscheidung operativer mathematischer Fähigkeiten können jedoch auch stoffgebietsspezifische Fähigkeiten differenziert werden. In Modell 3 werden in Anlehnung an die PISA-2003-Klassifikation vier interkorrelierte Fähigkeiten für die Stoffgebiete Arithmetik, Algebra, Geometrie und Stochastik differenziert (siehe auch Jordan u.a., 2006). Die Interkorrelationen zwischen den stoffgebietsspezifischen Fähigkeiten repräsentieren den Einfluss einer generellen mathematischen Fähigkeit.

Für die Spezifikation dieses Modells sprechen neben der empirischen Befundlage auch konzeptionelle und theoretische Gründe. So geht Vernon (1964) davon aus, dass die globale mathematische Fähigkeit hierarchisch stoffgebietsspezifischen Fähigkeiten übergeordnet ist. Carroll (1993) nimmt an, dass für die Fähigkeit „quantitatives Reasoning“ die Anwendung mathematischer Konzepte aus der Arithmetik, Algebra und Geometrie lösungsrelevant ist. Mayer (1985; vgl. Abschnitt 2.4) differenziert mehrere Arten mathematischen Wissens: schematisches, strategisches und prozedurales Wissen. Insbesondere beim schematischen Wissen ist davon auszugehen, dass dieses vornehmlich stoffgebietsspezifisch ist. So repräsentiert zum Beispiel die Formel zur Berechnung der Oberfläche einer Pyramide spezifisches schematisches Wissen aus der Geometrie. Es ist also denkbar, dass interindividuelle Unterschiede im stoffgebietsspezifischen Wissen (unabhängig davon, ob man Wissen deklarativ vs. prozedural oder im Sinne von Mayer differenziert) interindividuelle Unterschiede bei Aufgaben zur Messung mathematischer Schülerleistung erklären können.

Die Modelle 2 und 3 fokussieren auf die *mathematischen* Teilfähigkeiten mathematischer Schülerleistung. Die Fragen, wie man die Interkorrelationen der mathematischen Teilfähigkeiten im Rahmen hierarchischer Modelle spezifizieren kann und welche (nicht genuin mathematischen) *kognitiven* Fähigkeiten mathematischer Schülerleistung zu Grunde liegen, sind Gegenstand des Abschnitts 7.2.

Natürlich wäre zusätzlich zu den drei vorgestellten Modellen auch ein Modell interessant, bei dem simultan interindividuelle Unterschiede durch eine generelle mathematische Fähigkeit, operative Fähigkeiten und stoffgebietsspezifische Fähigkeiten (z.B. in Form eines Nested-Faktormodells) erklärt werden. Dieses Modell konnte aber mit den verfügbaren Daten nicht geprüft werden. Einerseits, weil die zwölf Zellen (4 Stoffgebiete  $\times$  3 Arten mathematischen Arbeitens) nicht mit genügend Mathematikaufgaben besetzt waren; andererseits, weil die Schüler nicht genügend Mathematikaufgaben der einzelnen Zellen bearbeiteten und somit die große Anzahl fehlender Werte keine stabile Schätzung der Modellparameter zuließ.

## 7.1.2 Methode

Zu den wichtigsten Ziele von PISA 2000 gehörte der Vergleich der mittleren Schülerleistungen unterschiedlicher Nationen und Bundesländer.

*Stichprobenziehung in PISA.* Bevor die Stichprobe beschrieben wird, die den Analysen in dieser Arbeit zu Grunde liegt, wird ein kurzer Überblick über die Stichprobenziehung und technischen Grundlagen der PISA-Studie gegeben. Die Stichprobenziehung bei PISA 2000 erfolgte mehrstufig (siehe z.B. Baumert & Artelt, 2003). In Deutschland wurden hierzu zunächst die Schulen nach Schulform, Bundesland und Schulgröße stratifiziert. Aus diesen Strata wurden im nächsten Schritt Schulen zufällig gezogen. Anschließend wurden (mit Ausnahme beruflicher Schulen) durch zufällige Ziehung innerhalb einer Schule zwei sich überlappende Schülerstichproben gebil-

det. Zur ersten Stichprobe gehörten 15-jährige Schüler unabhängig von der besuchten Klassenstufe. In der zweiten Stichprobe befanden sich Neuntklässler. Insgesamt nahmen in Deutschland 45.899 Schüler aus 1.466 Schulen an PISA 2000 teil (Baumert & Artelt, 2003).

*Beschreibung der Analysestichprobe.* Aus dem kompletten Datensatz von 45.899 Schülern wurde jedoch nur eine Teilstichprobe analysiert, die nachfolgend als *Analysestichprobe* bezeichnet wird. Die *Analysestichprobe* basiert auf der Anwendung von vier Auswahlkriterien.

- (a) Da der Erwerb mathematischen Wissens und mathematischer Fähigkeiten in hohem Maße von den schulischen Lerngelegenheiten abhängig ist, empfehlen Baumert, Stanat und Demmrich (2001, S. 34) eine Populationsdefinition nach Schulalter, wenn mathematische Kompetenzen untersucht werden. Daher wurden zur Beantwortung aller Forschungsfragen in dieser Arbeit nur die Daten von Schülern verwendet, die zum Testzeitpunkt die 9. Jahrgangsstufe besuchten. Das waren insgesamt 34.750 Schüler.
- (b) Die Daten für den internationalen und nationalen Vergleich wurden (meist) an zwei aufeinanderfolgenden Testtagen gesammelt. Allerdings nahmen nicht alle Schüler an beiden Testtagen teil. Durch das zweimalige Bearbeiten einer umfassenden Leistungstestbatterie kann einerseits die Testmotivation nachlassen, aber andererseits können auch leistungsförderliche Effekte aus der Testwiederholung resultieren (z.B. Kulik, Kulik & Bangert-Drowns, 1984). Um diese möglichen Varianzquellen konstant zu halten, wurden daher nur Schüler ausgewählt, die an beiden Testtagen teilgenommen hatten.
- (c) Im Mittelpunkt von PISA 2000 stand die Erfassung der Lesekompetenz der Schüler. Mathematik nahm eine untergeordnete Rolle ein. Aus dem ursprünglichen Datensatz der Neuntklässler wurden daher nur die Schüler ausgewählt, die eine größere Anzahl an Mathematikaufgaben bearbeitet hatten, damit reliable Indikatoren für die mathematische Schülerleistung gewonnen werden konnten. Einschlusskriterium war daher, dass für jeden Schüler für jeden der beiden Testtage mindestens drei Mathematikaufgaben vorliegen mussten, die als richtig oder falsch bewertet wurden. Damit wurde auch versucht, Schüler auszuwählen, die an beiden Testtagen ein Minimum an Mathematikaufgaben sachgemäß beantwortet hatten.
- (d) Weiterhin gingen die ausgewählten Schüler alle in Regelschulen. Sie bedurften also weder einer sonderpädagogischen Förderung, noch gingen sie auf die Berufsschule oder auf Waldorfschulen.

Trotz der angelegten Auswahlkriterien war die Analysestichprobe weitestgehend vergleichbar mit der ursprünglichen Stichprobe der Neuntklässler aus PISA 2000, wie eine „Ausfallanalyse“ hinsichtlich kognitiver, motivationaler und soziodemografischer Schülermerkmale zeigte (vgl. Anhang).

Insgesamt bestand die Analysestichprobe aus 29.386 Neuntklässlern (49,6% männlich) aus 1.301 Schulen. Das mittlere Alter lag bei 15,8 Jahren (SD 0,7 Jahre). Von den 29.386 Schülern besuchten 15,3 Prozent Hauptschulen, 24,1 Prozent Realschulen, 30,3 Prozent Gymnasien, 12,7 Prozent Gesamtschulen und 17,7 Prozent Schulen mit mehreren Bildungsgängen. Die Alters- und Geschlechterverteilung auf die einzelnen Schulformen ist in Tabelle 3 eingetragen. Zu beachten ist hierbei, dass mehr Jungen an die Hauptschule gingen und seltener das Gymnasium besuchten. Weiterhin war das Durchschnittsalter der Hauptschüler höher als das von Schülern an den anderen Schulformen.

Tabelle 3: Schüleranzahl, Geschlechterverteilung und mittleres Alter in Abhängigkeit der Schulform

Schulform	Schülerzahl	Anzahl an Schulen	Anteil der Jungen (%)	Mittleres Alter (SD)
Hauptschule	4.491	255	55,1	16,1 (0,7)
Realschule	7.077	303	50,2	15,9 (0,7)
Gymnasium	8.908	361	44,7	15,7 (0,6)
Gesamtschulen	3.719	162	51,3	15,8 (0,6)
Schulen mit mehreren Bildungsgängen	5.191	220	51,5	15,8 (0,6)
Insgesamt	29.386	1.301	49,6	15,8 (0,7)

Angaben basieren auf gültigen Werten der Schüler. Insgesamt lag für 60 Schüler keine Angabe zum Geschlecht vor und für 5 Schüler fehlte die Altersangabe.

*Testdesign und mathematischer Schülerleistungstest.* Um sowohl Aussagen über ein breites Kompetenzspektrum machen zu können als auch die Testdauer für jeden Schüler in zumutbaren Grenzen zu halten (Baumert u.a., 2001), wurden die PISA-Tests zur Erfassung der Schülerleistung in Lesen, Mathematik und in den Naturwissenschaften in Form eines Multi-Matrix-Sampling-Designs (Mislevy, Beaton, Kaplan & Sheehan, 1992) administriert. Die teilnehmenden Schüler erhielten jeweils an beiden Testtagen zufällig eines von neun Testheften zugewiesen (siehe für eine Übersicht über das Testdesign auch Anhang). Ein Teil der Aufgaben zur Erfassung der Schülerleistung in einem bestimmten Inhaltsgebiet kam in mehreren Testheften vor (Ankeritems), ein Teil der Aufgaben kam nur in einem bestimmten Testheft vor. Aufgrund der zufälligen Zuweisung sowie durch die verwendeten Ankeritems war es möglich, auf Basis von Item-Response-Modellen die Testleistungen aller Schüler auf einer gemeinsamen Metrik abzubilden.

An beiden Testtagen bearbeiteten die Schüler ungefähr für zwei Stunden Leistungstests (Schülerleistungstests, Tests zum schlussfolgernden Denken, Tests zum Problemlösen). Für die Erhebung der mathematischen Schülerleistung standen am ersten Testtag maximal 30 Minuten, am zweiten Testtag 25 Minuten zur Verfügung (Baumert u.a., 2001). Insgesamt wurden bei PISA 2000 117 verschiedene Mathematikitems eingesetzt. Diese wurden von der Expertengruppe Mathematik des PISA-Konsortiums nach mathematischen Stoffgebieten und den Arten mathematischen Arbeitens klassifiziert (Tab. 4; siehe auch Klieme, Neubrand u.a., 2001).<sup>16</sup> Von den 117 Mathematikaufgaben hatten 54 Items ein Multiple-Choice-Format, bei 63 Items konnte eine offene Antwort gegeben werden.

Durch das Multi-Matrix-Sampling-Design bearbeiteten nicht alle Schüler gleich viele Mathematikaufgaben. Aus Tabelle 5 ist beispielsweise ersichtlich, dass die Schüler im Mittel 30 Mathematikaufgaben bearbeiteten. Bei Anwendung des Einschlusskriteriums von mindestens drei (als richtig oder falsch) gewerteten Mathematikaufgaben pro Testtag (vgl. Auswahlkriterium c) hatten die Schüler in der Analysestichprobe minimal an beiden Testtagen zusammen 10 Aufgaben bearbeitet.

<sup>16</sup> Für die freundliche Überlassung des Datensatzes mit dieser Klassifikation danke ich Alexander Jordan.

Tabelle 4: Kreuzklassifikation aller Aufgaben des Mathematiktests aus PISA 2000 nach mathematischen Stoffgebieten und Arten mathematischen Arbeitens

	Algebra	Arithmetik	Geometrie	Stochastik	Summe
Technische Aufgaben	9	3	6	5	23
Rechnerisches Modellieren	8	19	12	8	47
Begriffliches Modellieren	16	7	14	10	47
Summe	33	29	32	23	117

Tabelle 5: Anzahl bearbeiteter Mathematikaufgaben pro Schüler

	Gesamt	Operative Teilfähigkeiten			Mathematische Stoffgebiete			
		Technische Aufgaben	Rechnerische Aufgaben	Begriffliche Modellierungsaufgaben	Arithmetikaufgaben	Algebraaufgaben	Geometrieaufgaben	Stochastikaufgaben
M	30	6	13	11	9	9	7	5
SD	7,7	1,5	3,6	4,2	1,5	2,2	3,3	3,4
Min.	10	1	5	2	0	2	1	1
P5	20	4	7	5	6	6	3	1
P25	22	5	9	7	8	7	4	2
P50	35	6	14	13	9	8	6	5
P75	37	7	16	15	10	11	10	8
P95	40	9	18	17	12	13	12	12
Max.	41	10	19	18	13	14	14	15

M = Mittelwert, SD = Standardabweichung, Min. = Minimum, P = Perzentil, Max. = Maximum. Die Anzahl bearbeiteter Mathematikaufgaben berechnete sich als die Anzahl richtig und falsch gewerteter Items (siehe Text).

Bei fast allen Mathematikaufgaben wurden nur zwei Scores unterschieden (0 für falsch, 1 für richtige Lösungen; dichotomes Scoringformat). Zehn Mathematikaufgaben hatten ein polytomes Scoringformat mit drei Abstufungen für die Richtigkeit der Lösung. Bei einer Aufgabe wurden vier Scores unterschieden. Fehlende Werte bei Mathematikaufgaben, die ein Schüler hätte bearbeiten können, wurden in dieser Arbeit als falsch gewertet (siehe auch Adams & Wu, 2002). Mathematikaufgaben, die ein Schüler nicht bearbeitet hat, weil ihm das jeweilige Testheft im Rahmen des Multi-Matrix-Sampling-Designs nicht vorgelegt wurde, wurden als fehlend (nicht administriert) codiert.

*Statistische Analysen.* Die Modelle 1 bis 3 wurden mit dem Programm ConQuest (Wu, Adams & Wilson, 1998) analysiert, das auch zur Skalierung der Schülerleistungstests in der PISA-Studie verwendet wurde. Bei Verwendung von ConQuest kann berücksichtigt werden, dass die Mathematikaufgaben im Rahmen eines Multi-Matrix-Sampling-Design administriert wurden und somit für alle Schüler zufällig fehlende Werte (*missing completely at random*, Rubin, 1987) bei einer Vielzahl

von Mathematikaufgaben vorlagen. Mit ConQuest ist es auch möglich, gleichzeitig Itemschwierigkeiten und Schwellenparameter für dichotome und polytome Items zu bestimmen (*Partial Credit Model*). Weiterhin ist es möglich Modellparameter für ein- und mehrdimensionale Rasch-Modelle zu schätzen. Wie beim herkömmlichen Modell der Faktorenanalyse wird bei mehrdimensionalen Rasch-Modellen davon ausgegangen, dass die latenten Variablen (die Faktoren) die Assoziationen zwischen den manifesten Variablen erklären können (Köller, Watermann & Baumert, 2001; siehe auch McDonald, 1997; Rost, 2004a). Die Verwendung eines uni- oder mehrdimensionalen Rasch-Modells wurde durch die bisherigen Modelltests der PISA-Mathematikaufgaben gestützt (Adams & Wu, 2002; Klieme, Neubrand u.a., 2001).

Die Modellparameter (Itemschwierigkeiten, Schwellenparameter, latente Varianzen/Kovarianzen) wurden in dieser Arbeit für alle Mathematikaufgaben frei geschätzt und *nicht* auf die Parameter aus der PISA-2000-Studie fixiert. Zwei Gründe sprachen hierfür: Erstens wurden für die PISA-Studie keine Itemparameter für Modelle mit mathematischen Stoffgebieten oder Arten mathematischen Arbeitens dokumentiert, da bei PISA 2000 in erster Linie nur der globale Mathematikscore von Interesse war. Zweitens war es durch die freie Schätzung der Modellparameter möglich, eine bestmögliche Modellanpassung an die Daten dieser Stichprobe zu erreichen. Somit wurde eine verzerrte Schätzung der latenten Korrelationen bei Fixierung auf die ursprünglichen Modellparameter von vornherein ausgeschlossen. Ein möglicher Nachteil bei dem gewählten Vorgehen war, dass die geschätzten Personenparameter dieser Studie nicht auf der originalen PISA-Metrik waren. Da jedoch kein direkter Vergleich mit den ursprünglichen PISA-Ergebnissen angestrebt wurde, schien dieser Nachteil nicht gravierend. Alle Modelle wurden identifiziert, indem der Mittelwert der latenten Variablen auf Null fixiert wurde (Wu u.a., 1998).

Wie auch bei den Modellanalysen in PISA 2000 (Adams & Wu, 2002, S. 101) wurden alle Analysen in dieser Arbeit *ungewichtet* durchgeführt. Damit geht jeder Schüler mit gleichem Gewicht (und nicht mit dem Gewicht, mit dem er die Population der Neuntklässler aus dem Jahr 2000 repräsentierte) zur Schätzung der Modellparameter ein.

Zur Beurteilung des Modell-Fits wurden drei Kriterien herangezogen:

- (a) Der globale Modell-Fit wurde anhand der  $\chi^2$ -*Goodness-of-Fit*-Werte beurteilt. Hierbei wird die Hypothese getestet, dass das spezifizierte Modell das „wahre“ Populationsmodell ist und die empirischen Daten „generiert“ hat. Jedoch ist vom  $\chi^2$ -*Goodness-of-Fit*-Test bekannt, dass bei Verwendung großer Stichproben selbst marginale Abweichungen des Modells von den empirischen Daten zu signifikanten Ergebnissen führen und damit alle Modelle verworfen werden (z.B. Hambleton, Swaminathan & Rogers, 1991). Ein Vorteil des  $\chi^2$ -*Goodness-of-Fit*-Tests ist jedoch, dass damit inferenzstatistische Aussagen über die Verbesserung des Modell-Fits bei zusätzlicher Spezifikation von Modellparametern möglich sind, wenn die zu vergleichenden Modelle ineinander genestet sind (Rost, 2004a). Hierzu wird die Differenz zwischen den  $\chi^2$ -Werten genesteter Modelle berechnet und auf Signifikanz beurteilt. Für alle inferenzstatistischen Tests wurde als Signifikanzniveau  $p < .05$  festgesetzt.
- (b) Unter anderem auch, um der Sensitivität des  $\chi^2$ -*Goodness-of-Fit*-Tests zu begegnen, ist ein weiteres Modellbewertungskriterium, wie gut ein bestimmtes Modell die empirischen Daten im Vergleich zu rivalisierenden Modellen approximiert (McDonald, 1997, 2000). Hierzu können die Informationskriterien (für die Berechnung und Interpretation siehe Rost, 2004a) dieser Modelle verglichen werden. Damit ist es auch möglich, Modellvergleiche durchzuführen, wenn die Modelle nicht ineinander genestet sind. Zur Berechnung des Informations-



kriteriums *Akaiikes Information Criterion* (AIC) werden Sparsamkeit des Modells (indiziert durch die Anzahl der geschätzten Modellparameter) und absolute Abweichung des Modells von den empirischen Daten (gemessen durch die Likelihood L der Daten „ $-2 \log(L)$ “, siehe Rost, 2004a) miteinander in Beziehung gesetzt: Wenn zum Beispiel zwei Modelle M1 und M2 die gleiche absolute Abweichung von den empirischen Daten aufweisen, in M1 jedoch mehr Modellparameter geschätzt wurden als in M2, dann wird aufgrund des AIC M2 favorisiert. Bei den Informationskriterien *Bayes Information Criterion* (BIC) und *Consistent AIC* (CAIC) wird noch zusätzlich die Anzahl der Modellparameter mit der (logarithmierten) Stichprobengröße<sup>17</sup> gewichtet (Rost, 2004a). Gemein ist allen drei Kriterien, dass kleinere Werte einen besseren Modell-Fit unter Berücksichtigung der Sparsamkeit des Modells anzeigen. Man betrachtet dann das Modell als beste Approximation, das die kleinsten Werte bei den Informationskriterien aufweist (Rost, 2004a) und theoretisch interpretierbar ist (siehe McDonald, 1997).

- (c) Ein weiteres Maß der Approximationsgüte eines Modells ist der Infit (*weighted mean-square error*, Wu, 1997) eines Einzelitems. Der Infit ist ein Maß dafür, wie gut die Lösungswahrscheinlichkeit, die durch das Modell vorhergesagt wird, mit der tatsächlichen Lösungshäufigkeit bei einem bestimmten Item übereinstimmt. Bei Modellkonformität eines Items nimmt der Infit den Wert 1 an. Infit-Werte eines Items zwischen 0,8 und 1,2 werden als akzeptable Abweichungen zwischen Modell und den beobachteten Daten betrachtet (Adams & Wu, 2002; Wang, Chen & Cheng, 2004).

### 7.1.3 Ergebnisse

Die Ergebnisse für den globalen Modell-Fit sind in Tabelle 6 eingetragen. Die signifikanten  $\chi^2$ -Werte aller drei Modelle legten den Schluss nahe, dass keines der drei Modelle das „wahre“ Modell war. Allerdings war es interessant zu fragen, ob die Spezifikation zusätzlicher latenter Variablen zu einer signifikanten Verbesserung des Modell-Fits führte. Diese Frage konnte mittels der  $\chi^2$ -Differenzentests beantwortet werden, wenn die zu vergleichenden Modelle ineinander genestet waren. Die Modelle 2 und 3 waren nicht ineinander genestet, jedoch waren jeweils die Modelle 2 und 3 in Modell 1 genestet: Wenn beispielsweise die latenten Korrelationen zwischen den stoffgebietspezifischen Faktoren auf „1,0“ fixiert werden und somit die Identität der Faktoren impliziert wird, entspricht dies einem einfaktoriellen Modell. Sowohl die  $\chi^2$ -Differenz ( $\Delta\chi^2$ ) für den Vergleich von Modell 2 mit Modell 1 ( $\Delta\chi^2 = 977$ ,  $df = 5$ ) als auch für den Vergleich von Modell 3 mit Modell 1 ( $\Delta\chi^2 = 3.017$ ,  $df = 9$ ) waren signifikant. Beide Modelle passen also aus inferenzstatistischer Perspektive besser als Modell 1.

Diese Schlussfolgerung wurde auch durch die Analyse der Informationskriterien gestützt: Modell 2 als auch Modell 3 hatten bei allen drei Kriterien niedrigere Werte als Modell 1. Im Vergleich von Modell 2 mit Modell 3 indizierten alle drei Informationskriterien, Modell 3 mit den stoffgebietspezifischen Fähigkeiten zu favorisieren.

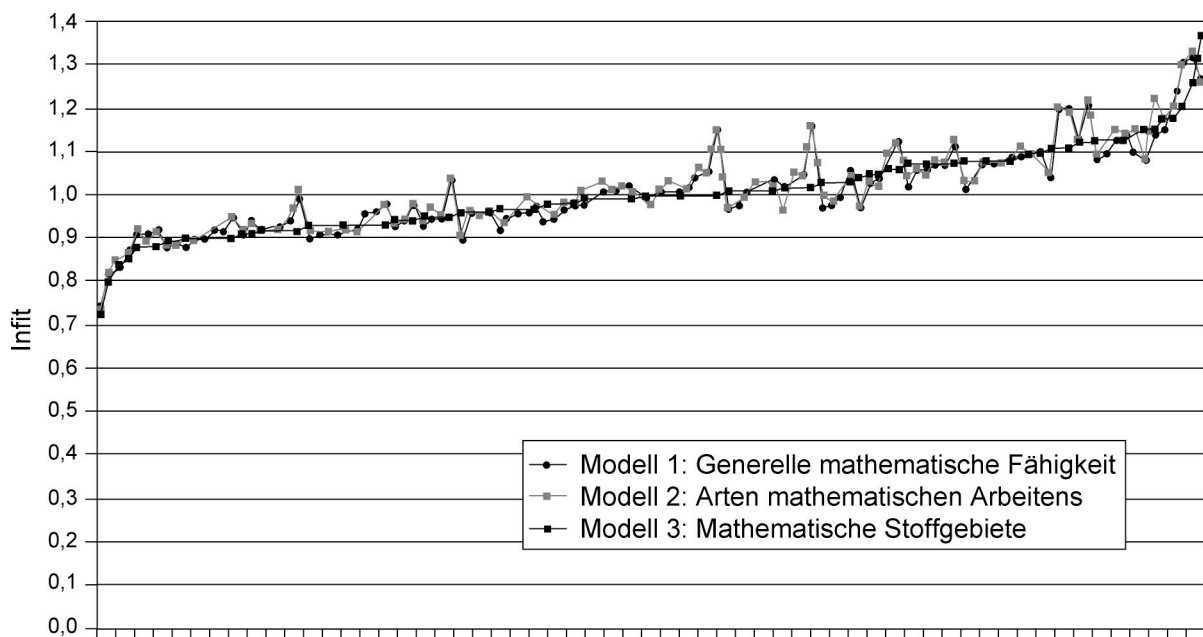
---

<sup>17</sup> Bei der Berechnung von BIC und CAIC wurde als Stichprobengröße ein N von 29.386 verwendet und kein für den Designeffekt, der aus der mehrstufigen Stichprobenziehung resultierte, korrigiertes N (vgl. Snijders & Bosker, 1999). Hierfür sprach in erster Linie, dass der Designeffekt sich bei den spezifizierten latenten Variablen unterscheiden konnte, aber bei allen Modellen dasselbe N eingehen sollte.

Tabelle 6: Globaler Modell-Fit:  $\chi^2$ -*Goodness-of-Fit*-Test und Informationskriterien

Modell	$\chi^2$	df	AIC	BIC	CAIC
Modell 1: Generelle mathematische Fähigkeit	961.785	132	962.049	962.374	962.506
Modell 2: Arten mathematischen Arbeitens	960.808	137	961.082	961.420	961.557
Modell 3: Mathematische Stoffgebiete	957.791	141	<b>958.073</b>	<b>958.421</b>	<b>958.562</b>

df = Freiheitsgrade. Die kleinsten Werte von AIC = *Akaike's Information Criterion*, BIC = *Bayes Information Criterion* und CAIC = *Consistent AIC* sind fett gedruckt.



Die Werte sind aufsteigend für Modell 3 sortiert.

Abbildung 15: Infit-Werte aller 117 Mathematikaufgaben aus dem PISA-2000-Test

Auch eine Inspektion der Infit-Werte legte nahe, Modell 3 gegenüber Modell 2 den Vorzug zu geben: In Modell 1 und in Modell 2 hatten insgesamt sechs Mathematikaufgaben und in Modell 3 drei Mathematikaufgaben Infit-Werte außerhalb des kritischen Ranges von 0,8 bis 1,2. Abbildung 15 veranschaulicht diesen Sachverhalt. Darin sind die Infit-Werte aller 117 Mathematikaufgaben für alle drei Modelle dargestellt.

Alle globalen Modell-Fit-Indizes favorisierten Modell 3 und damit die Unterscheidung stoffgebietspezifischer mathematischer Fähigkeiten. Allerdings deuteten die Infit-Werte von drei (Modell 3) bzw. von sechs Mathematikaufgaben (Modelle 1 und 2) außerhalb des kritischen Ranges auf eine mögliche Unvereinbarkeit dieser Aufgaben mit einem Raschmodell hin. Um die Robustheit der Schätzung der latenten Korrelationen gegenüber diesen Modellabweichungen zu überprüfen, wurden die latenten Korrelationen in den Modellen 2 und 3 jeweils mit und ohne die Mathematikaufgaben berechnet, die außerhalb des kritischen Infit-Ranges lagen.

Tabelle 7: Modell 2: Latente Korrelationen der operativen mathematischen Fähigkeiten/Arten mathematischen Arbeitens

	Technische Fertigkeiten	Rechnerisches Modellieren	Begriffliches Modellieren
Technische Fertigkeiten	1,20	0,85	0,87
Rechnerisches Modellieren	0,85	1,07	0,94
Begriffliches Modellieren	0,87	0,94	1,09

Unterhalb der Diagonalen sind die latenten Interkorrelationen (in der Diagonalen die latenten Standardabweichungen) berechnet auf der Basis aller 117 Mathematikaufgaben eingetragen. Oberhalb der Diagonalen stehen die latenten Interkorrelationen nur für die 111 Mathematikaufgaben mit akzeptablen Infit-Werten.

Tabelle 8: Modell 3: Latente Korrelationen der stoffgebietspezifischen Fähigkeiten

	Arithmetik	Algebra	Geometrie	Stochastik
Arithmetik	1,30	0,89	0,86	0,80
Algebra	0,88	1,14	0,89	0,80
Geometrie	0,86	0,87	1,10	0,78
Stochastik	0,80	0,79	0,78	0,96

Unterhalb der Diagonalen sind die latenten Interkorrelationen (in der Diagonalen die latenten Standardabweichungen) berechnet auf der Basis aller 117 Mathematikaufgaben eingetragen. Oberhalb der Diagonalen stehen die latenten Interkorrelationen nur für die 114 Mathematikaufgaben mit akzeptablen Infit-Werten.

In den Tabellen 7 und 8 sind die latenten Korrelationen der operativen mathematischen Fähigkeiten (Arten mathematischen Arbeitens) und der stoffgebietspezifischen Fähigkeiten eingetragen. Die latenten Interkorrelationen in Modell 2 waren hoch, jedoch von 1,0 verschieden: Sie lagen zwischen 0,85 (technische Fertigkeiten und rechnerisches Modellieren) und 0,94 (rechnerisches Modellieren und begriffliches Modellieren). In Modell 3 resultierten ebenfalls hohe latente, aber von 1,0 verschiedene Korrelationen: Die Korrelationskoeffizienten lagen zwischen 0,78 (Stochastik und Geometrie) und 0,88 (Algebra und Arithmetik).

Die latenten Korrelationen waren auch robust gegenüber der zu Grunde liegenden Auswahl (in Bezug auf die Infit-Werte) an Mathematikaufgaben. So resultierten (auf zwei Nachkommastellen gerundet) keine Unterschiede für die Korrelationen in Modell 2. In Modell 3 betrug die größte Differenz zwischen korrespondierenden Korrelationen .02.

#### 7.1.4 Diskussion

Die beiden wichtigsten Ergebnisse der vorangegangenen Analysen können folgendermaßen zusammenfasst werden.

- Angesichts des besseren Modell-Fits wurde Modell 3 mit stoffgebietspezifischen Fähigkeiten favorisiert. Dieses Modell konnte die empirischen Relationen besser erklären als Modell 1, das

nur eine generelle mathematische Fähigkeit vorsah, oder Modell 2, in dem operative Teilfähigkeiten nach den Arten mathematischen Arbeitens differenziert wurden.

- Unabhängig davon, ob spezifische mathematische Fähigkeiten nach operativen oder inhaltlichen Gesichtspunkten unterschieden wurden, resultierten hohe latente Interkorrelationen dieser Fähigkeiten.

Wie sind die stoffgebietsspezifischen Fähigkeiten zu interpretieren? Sie repräsentieren sensu Carroll (1993; vgl. Abschnitt 2.2) die Fähigkeit zur Anwendung von stoffgebietsspezifischem Mathematikwissen. Welche Wissensarten (z.B. deklarativ, prozedural oder konzeptuell) hierbei den größten Anteil der interindividuellen Unterschiede erklären, konnte auf Grundlage des verwendeten Aufgabenmaterials jedoch nicht beantwortet werden. Fest steht aber, dass alle Aufgaben des PISA-Mathematiktests das Generieren einer Lösung erforderten und somit prozedurale Aspekte betonten. Keine einzige Aufgabe erforderte ausschließlich die Wiedergabe von deklarativem mathematischem Faktenwissen.

Angesichts der operativen Differenzierung mathematischer Fähigkeit in den Strukturmodellen kognitiver Fähigkeit (vgl. Abschnitt 4.2) kann spekuliert werden, ob diese Trennung nicht ein „Artefakt“ des verwendeten Aufgabenmaterials ist. In der Auflistung von operativ orientierten mathematischen Fähigkeiten (Carroll, 1993; Horn & Noll, 1997) fehlen meist Aufgaben aus den anderen mathematischen Stoffgebieten, insbesondere Geometrie und Stochastik. Betrachtet man beispielsweise die Markieraufgaben von Rechenfertigkeit, so stammen diese meist aus der Arithmetik. Auch viele Aufgaben zur Erfassung von quantitativem Reasoning können der Arithmetik zugeordnet werden. Analysiert man solche Aufgabenbatterien mit exploratorischen Faktoranalysen, führt dies zur Extraktion von Faktoren, die operative mathematische Fähigkeiten repräsentieren. Dies jedoch nur, weil die mathematischen Stoffgebiete nicht vollständig repräsentiert sind, oder um mit Anastasi zu sprechen: „What emerges from a factor analysis depends on what variables are included in the correlation matrix.“ (Anastasi, 1986, S. 196)

Möglicherweise sind also die operativen mathematischen Teilfähigkeiten im Rahmen der Strukturforschung kognitiver Fähigkeiten nur deshalb gefunden worden, weil die anderen mathematischen Stoffgebiete nicht mit genügend Aufgaben repräsentiert waren. Da mit Blick auf Arithmetik in den Strukturmodellen kognitiver Fähigkeiten sowohl fertigungs- als auch problemorientierte Fähigkeiten unterschieden werden, wirft dies die Frage auf, ob die stoffgebietsspezifischen Fähigkeiten nicht selbst wiederum hierarchisch den operativen Teilfähigkeiten übergeordnet sind.<sup>18</sup> Möchte man also dem Ziel nahe kommen, eine Taxonomie kognitiver Fähigkeiten zu erstellen (Gustafsson & Undheim, 1996), ist bei Zusammenstellung der Testbatterien zu beachten, dass operative Gesichtspunkte mathematischer Fähigkeiten, aber auch die mathematischen Stoffgebiete ausgewogen berücksichtigt werden.

Unabhängig von diesen Spekulationen zeigten die hohen latenten Korrelationen in den Modellen 2 und 3, dass technische Fertigkeiten, rechnerisches Modellieren und begriffliches Modellieren, aber auch stoffgebietsspezifische Fähigkeiten nur eingeschränkt unterschieden werden konnten. Welche Erklärungen bieten sich hierfür an?

---

<sup>18</sup> Diese Frage kann jedoch auf der Grundlage des PISA-Mathematiktests aus dem Jahr 2000 nicht befriedigend beantwortet werden, wie in Abschnitt 7.1.1 bereits erörtert wurde.

Ein Grund liegt in der Testkonstruktion des PISA-Mathematiktests verborgen. Der Test wurde so konstruiert, dass die ausgewählten Mathematikaufgaben bei 15-jährigen Jugendlichen mit einem eindimensionalen Raschmodell vereinbar sind. Mathematikaufgaben wurden dabei nach Arten mathematischen Arbeitens oder mathematischen Stoffgebieten klassifiziert, und es wurde darauf geachtet, dass diese in einem ausgewogenen Verhältnis in der Feldtestphase enthalten waren. Allerdings wurde nicht versucht, spezifische Varianzen der Arten mathematischen Arbeitens oder der mathematischen Stoffgebiete zu maximieren, um damit die latenten Interkorrelationen zu verringern.

Allerdings sind die hohen latenten Interkorrelationen zwischen operativen oder inhaltlichen mathematischen Fähigkeiten kein genuines Charakteristikum des PISA-Mathematiktests. Auch in den anderen *Large-Scale*-Studien, die in Abschnitt 7.1.1 referiert wurden, waren die Korrelationen zwischen den operativen oder stoffgebietsspezifischen Fähigkeiten hoch. Insgesamt gesehen scheint es – zumindest im Rahmen von *Large-Scale*-Studien – ein prinzipielles Problem zu sein, verschiedene mathematische Prozesse oder Wissensarten psychometrisch getrennt zu erfassen (siehe hierzu auch Abschnitt 2.6 und 10.3.1).

Der Befund hoher latenter Interkorrelationen wird jedoch nicht nur durch Einflüsse der Testkonstruktion erklärt. Auch theoretische Überlegungen unterstützen dieses Ergebnis:

- (a) Mathematische Fähigkeiten entwickeln sich kumulativ: Höhere mathematische Fähigkeiten bauen auf weniger komplexen einfacheren mathematischen Fähigkeiten auf (Geary, 1995). Man kann mit den Autoren des PISA-Mathematiktests (Blum u.a., 2004; Klieme, Neubrand u.a., 2001) argumentieren, dass technische Fertigkeiten weniger komplexe Fähigkeiten repräsentieren und rechnerisches und begriffliches Modellieren höhere mathematische Fähigkeiten darstellen. Bei Schülern, bei denen stoffgebietsübergreifend technische Fertigkeiten weiter entwickelt sind, sollten somit stoffgebietsübergreifend komplexe Fähigkeiten (rechnerisches und begriffliches Modellieren) ebenfalls weiterentwickelt sein. Dies würde dazu führen, dass bei einer querschnittlichen Analyse mathematische Fähigkeiten hoch interkorrelieren, unabhängig davon, ob man mathematische Fähigkeiten operativ oder inhaltlich differenziert.
- (b) Ein weiterer Grund für die Interkorrelation mathematischer Fähigkeiten sind Unterschiede in den schulischen Lernangeboten. Wie in Abschnitt 5.3 und 5.4 herausgearbeitet wurde, unterscheiden sich Schulen und Schulklassen in der Qualität des Mathematikunterrichts. Wenn diese Unterschiede zeitlich stabil sind, können sie zur Faktorintegration spezifischer mathematischer Fähigkeiten führen (vgl. Abschnitt 4.3). Dies kann letztlich eine Differenzierung der mathematikspezifischen Fähigkeit  $M'$  (siehe Abb. 16, Modell 7a) implizieren. Der Frage, wie Lernumgebungen und die Heterogenität der mathematikspezifischen Fähigkeit zusammenhängen, wird im Rahmen der zweiten Forschungsfrage nachgegangen und an dieser Stelle nicht weiter elaboriert.
- (c) Unabhängig von den Lernumgebungen spiegeln die latenten Interkorrelationen auch den Einfluss einer Fähigkeit wider, die zum Lösen aller Aufgaben des PISA-Mathematiktests notwendig ist: Rechnerische und begriffliche Modellierungsaufgaben erfordern das Lösen von Problemen. Die sehr hohen Korrelationen von technischen Fertigkeiten und den beiden Fähigkeiten zum mathematischen Modellieren können so interpretiert werden, dass technische Aufgaben für viele Schüler nicht nur den Abruf einer automatisierten Fertigkeit, sondern auch Problemlöseprozesse erforderten. Angesichts der vorliegenden Ergebnisse wäre daher ein alternativer Begriff für technische Aufgaben „Routineprobleme“ (siehe Mayer &

Hegarty, 1996, S. 32). Dies zusammen mit der Tatsache, dass im PISA-Mathematiktest problemorientierte Aufgaben dominieren (von 117 Mathematikaufgaben erfassten nur 23 Aufgaben technische Fertigkeiten), legt den Schluss nahe, dass die Fähigkeit zum mathematischen Problemlösen am Zustandekommen dieser Interkorrelationen bedeutsam beteiligt ist. Vor dem Hintergrund der Ergebnisse aus Abschnitt 7.2 kann diese Interpretation in Abschnitt 7.3 noch weiter gestützt und theoretisch geschärft werden.

- (d) Es stellt sich aber auch die Frage, ob die latenten Interkorrelationen zwischen den mathematischen Fähigkeiten nur durch *eine* (generelle) mathematische Fähigkeit oder auch noch zusätzlich durch andere kognitive Fähigkeiten erklärt werden können. Diese Frage ist Gegenstand des nächsten Kapitels.

## 7.2 Mathematische Fähigkeiten und (nicht mathematische) kognitive Fähigkeiten

### 7.2.1 Empirische Befundlage und zu prüfende Strukturmodelle

Dieser Abschnitt befasst sich mit der Frage, welche kognitiven Fähigkeiten interindividuelle Unterschiede bei (aggregierten) Maßen mathematischer Schülerleistung erklären können. Nachfolgend werden hierzu einige Studien vorgestellt, die nicht mehr wie im Abschnitt 7.1 auf der Ebene von Einzelitems ansetzen, sondern auf der Ebene spezifischer mathematischer Fähigkeiten.

*Treumanns Literaturübersicht.* In einem Gutachten für die Bildungskommission des deutschen Bildungsrats zeigte Treumann (1974) in einer Literaturübersicht, welche kognitiven Fähigkeiten interindividuellen Unterschieden bei Maßen mathematischer Schülerleistung zu Grunde liegen. Er begutachtete hierfür fast alle relevanten Studien, die bis zum Jahr 1974 durchgeführt wurden. Von Interesse für die vorliegende Arbeit sind die Faktorladungen, die angeben, wie stark eine bestimmte kognitive Fähigkeit die Leistung bei Maßen mathematischer Schülerleistung beeinflusst. Treumann fasste die Ladungen über alle von ihm begutachteten Studien zusammen und teilte dabei die Studien danach auf, welche faktoranalytische Methode verwendet wurde. In Tabelle 9 sind die von Treumann berichteten Mediane der standardisierten Ladungen für Indikatoren stoffgebietsspezifischer Fähigkeiten eingetragen.

Das wichtigste Ergebnis aus Treumanns Literaturübersicht lässt erkennen (siehe Tab. 9), dass unabhängig von der verwendeten faktoranalytischen Technik interindividuelle Unterschiede bei stoffgebietsspezifischen Fähigkeiten durch eine globalere mathematische Fähigkeit (M oder QR<sup>19</sup>) und durch die allgemeine kognitive Fähigkeit erklärt werden konnten. Die Erklärungskraft (gemessen an den Faktorladungen) von Rechenfertigkeit, verbaler Fähigkeit, visuell-räumlicher Fähigkeit war deutlich geringer und schien zudem von der gewählten faktoranalytischen Technik abzuhängen.

Letzteres war insbesondere auch der Fall bei quantitativem Reasoning. Hierfür dokumentierte Treumann bei der hierarchischen Gruppenfaktorentechnik keine Faktorladungen. Ein möglicher Grund ist, dass der Varianzanteil von schlussfolgerndem Denken bereits von allgemeiner kogni-

---

<sup>19</sup> Diesen Faktor bezeichnet Treumann (1974) als allgemeines Denken. Markieraufgaben wie auch Faktordefinition sind dabei nahezu identisch mit dem Faktor quantitatives Reasoning im Modell von Carroll (1993). Zur Vereinheitlichung des Sprachgebrauchs wird dieser Faktor daher als quantitatives Reasoning bezeichnet.

tiver Fähigkeit repräsentiert wurde (vgl. Gustafsson, 1984). Somit würde der Faktor M, der bei der hierarchischen Gruppenfaktorentechnik mathematisches Wissen repräsentierte, den mathematikspezifischen Varianzanteil von quantitativem Reasoning (wie es in einem Modell mit interkorrelierten Gruppenfaktoren konzeptionalisiert wird) darstellen.

Tabelle 9: Mediane der Faktorladungen von Schülerleistungstests in 30 Studien (nach Treumann, 1974, S. 351, Tab. IV)

Stoffgebiet	N	V	V-R	M	QR	g
<i>Interkorrelierte Gruppenfaktoren</i>						
Arithmetik	.29	.26	.11	.46	.48	.56 <sup>a</sup>
Algebra	.15	.21	.23	.67	.38	.62 <sup>a</sup>
Geometrie	.05	.08	.18	.50	.37	.55 <sup>a</sup>
<i>Hierarchische Gruppenfaktorentechnik</i>						
Arithmetik	.11	.00	-.01	.24	x	.73
Algebra	.05	.09	.02	.19	x	.76
Geometrie	-.10	.12	.18	.20	x	.68

N = Rechenfertigkeit, V = verbale Fähigkeit, V-R = visuell-räumliche Fähigkeit, M = mathematisches Wissen, QR = quantitatives Reasoning, g = allgemeine kognitive Fähigkeit, x = keine Angabe in Treumann (1974).

<sup>a</sup> Von Treumann geschätzte Faktorladung (Wurzel aus der Summe der quadrierten Faktorladungen über alle als Reasoning identifizierte Faktoren hinweg).

*Aktuellere Einzelstudien.* Treumann fasste die bis 1974 publizierten Befunde bündig zusammen. Es stellt sich damit die Frage, ob seine Ergebnisse auch in jüngerer Zeit repliziert werden konnten. Nachfolgend werden aktuellere Einzelstudien aufgelistet, die Maße mathematischer Schülerleistung oder mathematische Fähigkeiten und (nicht genuin mathematische) Fähigkeiten analysierten.

Gustafsson konnte mittels konfirmatorischer Faktorenanalysen unter Verwendung von hierarchischen Faktormodellen (Gustafsson, 1984) und Nested-Faktormodellen (Gustafsson & Balke, 1993; Gustafsson, 1994) zeigen, dass mehrere kognitive Fähigkeiten interindividuelle Unterschiede bei Maßen mathematischer Schülerleistung erklären können. Studienübergreifend erklärte die fluide Fähigkeit, die Gustafsson als äquivalent zu allgemeiner kognitiver Fähigkeit betrachtet, den größten Anteil der Varianz bei den Maßen mathematischer Schülerleistung. Zusätzlich erklärten in allen drei Studien eine generelle mathematische Fähigkeit und in den Studien aus den Jahren 1984 und 1993 eine verbale Fähigkeit (markiert durch Schülerleistungstests für Schwedisch und Englisch) kleinere Varianzanteile.

Im Rahmen von zwei Studien analysierte McGrew den Zusammenhang von quantitativem Reasoning und kognitiven Fähigkeiten. Bei 15- bis 16-jährigen amerikanischen Jugendlichen gehörten zu den bedeutsamsten Prädiktoren mathematischer Fähigkeiten die Informationsverarbeitungsgeschwindigkeit, eine verbale Fähigkeit und die fluide Fähigkeit (McGrew & Hessler, 1995, S. 27). In einer weiteren Studie analysierte McGrew eine altersgruppenübergreifende, amerikanische Stichprobe. Das Maß für quantitatives Reasoning wurde bei einer konfirmatorischen Faktoren-

analyse bedeutsam von einer generellen mathematischen Fähigkeit, einer verbalen Fähigkeit und einer visuell-räumlichen Fähigkeit beeinflusst (McGrew, 1997, S. 178).

Bickley, Keith und Wolfle (1995) untersuchten mittels konfirmatorischer Faktorenanalysen den Zusammenhang von Maßen mathematischer Fähigkeiten mit weiteren Maßen kognitiver Fähigkeiten. Drei Ergebnisse sind für die vorliegende Arbeit von Bedeutung. Erstens konnte eine generelle mathematische Fähigkeit von anderen Fähigkeiten faktoriell unterschieden werden. Zweitens hatte diese generelle mathematische Fähigkeit sehr hohe Ladungen auf einem Faktor, der die allgemeine kognitive Fähigkeit repräsentierte. Drittens konnte dieser Befund altersgruppenübergreifend für die Altersspanne von 6 bis 79 Jahre repliziert werden.

Bei der deutschen PISA-2000-Stichprobe der 15-jährige Jugendlicher erklärten die fluide Fähigkeit, eine generelle verbale Fähigkeit (gemessen durch den PISA-Lesetest) bedeutsam interindividuelle Unterschiede in der mathematischen Schülerleistung (Klieme, Neubrand u.a., 2001). Dieser Befund konnte bei PISA 2003 für die Stichprobe der 15-jährigen Jugendlichen repliziert werden (Leutner, Klieme, Meyer & Wirth, 2004).

Mittels konfirmatorischer Faktorenanalysen analysierten Brunner und Süß (2005) Daten von deutschsprachigen Personen, die den Berliner Intelligenzstrukturtest bearbeitet hatten. Das wichtigste Ergebnis für die vorliegende Arbeit war, dass interindividuelle Unterschiede bei mathematischen Aufgaben durch die jeweilige operative Facette, eine generelle mathematische Fähigkeit (dies entspricht im BIS-Modell numerischer Fähigkeit) und vor allem durch die allgemeine kognitive Fähigkeit erklärt wurden.

*Zusammenfassung.* Zusammenfassend kann man festhalten, dass Treumanns Literaturübersicht und die referierten Befunde aktuellerer Studien ein recht eindeutiges Bild ergeben, welche kognitiven Fähigkeiten interindividuelle Unterschiede bei Maßen mathematischer Schülerleistung erklären können. Studienübergreifend waren das die allgemeine kognitive Fähigkeit und eine generelle mathematische Fähigkeit. In einigen weiteren Arbeiten hatte auch die verbale Fähigkeit einen positiven Einfluss auf die Leistung. Die Zusammenhänge zwischen spezifischen mathematischen Fähigkeitsmaßen und Informationsverarbeitungsgeschwindigkeit (McGrew & Hessler, 1995), Visualisierung (McGrew, 1997) und operativen Fähigkeiten (Brunner & Süß, 2005) waren studienübergreifend nicht replizierbar. Gründe hierfür sind im komplexen Zusammenspiel der verwendeten Aufgabenbatterie, der statistischen Analysemethoden (Regression, interkorrelierte Gruppenfaktoren, hierarchische Gruppenfaktorentechnik oder Nested-Faktormodelle) und der Stichprobenzusammensetzung zu suchen.

Diese Schlussfolgerung deckt sich auch gut mit den Ableitungen, die anhand der Strukturmodelle kognitiver Fähigkeiten gemacht wurden: Vernon (1964), Carroll (1993) und Jäger und Kollegen (1997) gehen von einer globaleren mathematischen Fähigkeit aus und nehmen an, dass die Leistungen bei Maßen kognitiver Fähigkeiten multipel bedingt sind. Folgt man dieser Annahme, erklären die allgemeine kognitive Fähigkeit und eine globalere mathematische Fähigkeit zusätzlich zu spezifischen mathematischen Fähigkeiten interindividuelle Unterschiede bei Maßen mathematischer Schülerleistung. Weiterhin geht Vernon davon aus, dass interindividuelle Unterschiede bei verbaler und numerischer Fähigkeit durch einen verbal-numerischen Bildungsfaktor erklärt werden können. Dies korrespondiert gut mit dem empirischen Befund, dass Maße mathematischer Schülerleistung zusätzlich durch eine verbale Fähigkeit beeinflusst werden (siehe auch Carroll, 1996).



Im Unterschied zu den obigen drei Autoren gibt es nach Horn und Noll (1997) keine allgemeine kognitive Fähigkeit. Die (reliable) Varianz bei Maßen mathematischer Schülerleistung setzt sich somit nur aus einer generellen mathematischen Fähigkeit und spezifischen mathematischen Fähigkeiten zusammen.

*Zu prüfende Strukturmodelle.* Nachfolgend wird der Frage nachgegangen, welche kognitiven Fähigkeiten interindividuelle Unterschiede bei Maßen mathematischer Schülerleistung erklären können. Vor dem referierten empirischen und theoretischen Hintergrund (vgl. Abschnitt 4.2) wird dabei besonderes Augenmerk auf eine globalere mathematische Fähigkeit, die allgemeine kognitive Fähigkeit und die verbale Fähigkeit gelegt. Da in der psychometrisch-orientierten Forschung kognitive Fähigkeiten in der Regel faktoranalytisch abgeleitet werden, wird diese Frage mit Hilfe von Strukturmodellen untersucht (siehe Abb. 16). Auf Grundlage der Ergebnisse zu den spezifischen mathematischen Fähigkeiten in Abschnitt 7.1, werden nachfolgend nur stoffgebietsspezifische Fähigkeiten betrachtet. Die Modelle werden dabei fortlaufend durchnummeriert, um die Modelle aus Abschnitt 7.1 klar von den Modellen aus Abschnitt 7.2 zu trennen.

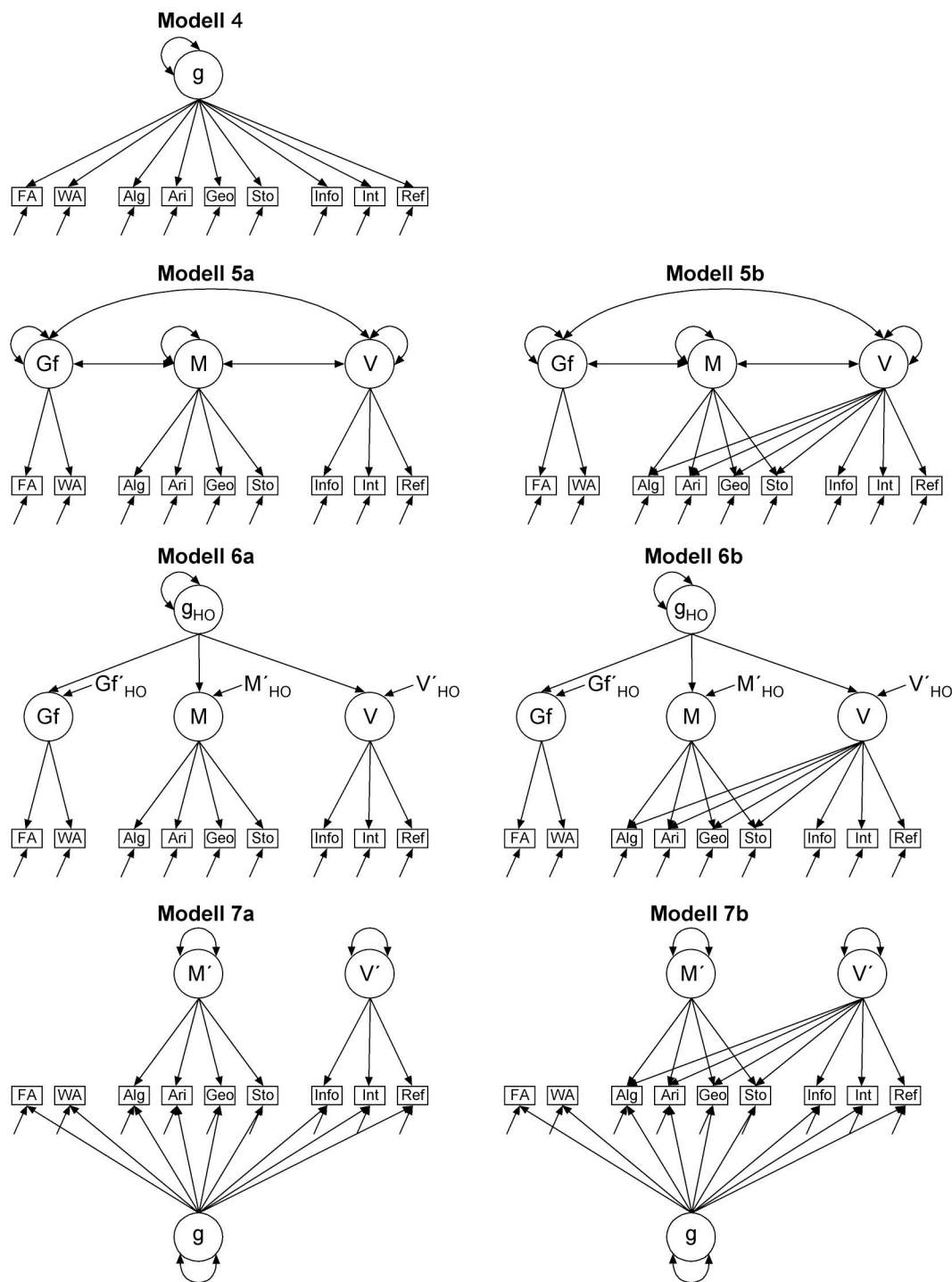
*Modell 4: Allgemeine kognitive Fähigkeit.* Mit Spearman kann man davon ausgehen, dass die allgemeine kognitive Fähigkeit ( $g$ ) und stoffgebietsspezifische Fähigkeiten die interindividuellen Unterschiede bei den Maßen mathematischer Schülerleistung erklären können. Stoffgebietsspezifische Fähigkeiten werden durch die Residualterme der manifesten Variablen dargestellt. Mit diesem einfaktoriellen Modell wird die Annahme getestet, dass die Interkorrelationen zwischen den stoffgebietsspezifischen Fähigkeiten nur durch die allgemeine kognitive Fähigkeit erklärt werden.

Gleichzeitig stellt der Vergleich zu Modell 4 einen empirischen Bezugspunkt dafür dar, ob die nachfolgend dargestellten Modelle die empirischen Relationen zwischen den Maßen mathematischer Schülerleistung besser als Modell 4 approximieren können.

*Modell 5a: Standardmodell (fluide und kristalline Fähigkeiten).* Modell 5a folgt den Annahmen der Theorie fluider und kristalliner Fähigkeiten (Horn & Noll, 1997). Nicht zuletzt aufgrund der Popularität des Gf-Gc-Modells (Gustafsson & Undheim, 1996) ist meines Wissens Modell 5a das am weitesten gebräuchliche Modell im pädagogisch-psychologischen Forschungskontext (siehe z.B. die Untersuchungen von Renkl & Stern, 1994; Weinert & Helmke, 1995a, 1998) und wird daher nachfolgend als *Standardmodell* bezeichnet.

Die fluide Fähigkeit (Gf), eine generelle verbale Fähigkeit (V) und eine generelle mathematische Fähigkeit (M) werden konzeptuell voneinander unterschieden. Die manifesten Maße mathematischer Schülerleistung werden in diesem Modell als eine Funktion stoffgebietsspezifischer Fähigkeiten (repräsentiert durch die Residualterme) und der generellen mathematischen Fähigkeit betrachtet.

Wie sind die latenten Korrelationen zu interpretieren? Horn und Noll sagen explizit, dass die latenten Korrelationen nicht den Einfluss der allgemeinen kognitiven Fähigkeit widerspiegeln (1997, S. 68), machen aber keine Aussage, wie die latenten Interkorrelationen sonst interpretiert werden können. Eine mögliche Deutung kann durch die Investmenttheorie (vgl. Abschnitt 5.2; Cattell, 1987) gegeben werden. Die latenten Korrelationen repräsentieren somit die Varianzanteile, die auf die investierte fluide Fähigkeit zurückgehen.



g = allgemeine kognitive Fähigkeit, Gf = fluide Fähigkeit, M = generelle mathematische Fähigkeit, V = verbale Fähigkeit,  $g_{HO}$  = allgemeine kognitive Fähigkeit,  $Gf'_{HO}$  = residualisierte fluide Fähigkeit,  $M'_{HO}$  = residualisierte mathematische Fähigkeit,  $V'_{HO}$  = residualisierte verbale Fähigkeit,  $M'$  = mathematikspezifische Fähigkeit,  $V'$  = spezifische verbale Fähigkeit, FA = Figurenanalogien, WA = Wortanalogien, Alg = Algebra, Ari = Arithmetik, Geo = Geometrie, Sto = Stochastik, Info = Informationen ermitteln, Int = textbezogenes Interpretieren, Ref = Reflektieren und Bewerten. Für eine Beschreibung der manifesten Variablen siehe Abschnitt 7.2.2.

Abbildung 16: Modelle zur Überprüfung des Zusammenhangs von mathematischen Fähigkeiten und (nicht genuin mathematischen) kognitiven Fähigkeiten

*Modell 5b: Standardmodell-verbal.* Dieses Modell ist konzeptuell identisch mit dem Standardmodell. Einziger Unterschied ist, dass interindividuelle Unterschiede bei den Maßen mathematischer Schülerleistung zusätzlich durch eine generelle verbale Fähigkeit erklärt werden.

*Modell 6a: „Higher-Order“-Faktorenmodell.* Wie im Standardmodell werden die manifesten Maße mathematischer Schülerleistung in diesem Modell als eine Funktion stoffgebietspezifischer Fähigkeiten und der generellen mathematischen Fähigkeit (M) betrachtet. Im Gegensatz zum Standardmodell wird im *Higher-Order*-Modell explizit angenommen, dass die allgemeine kognitive Fähigkeit ( $g_{HO}$ ) die latenten Interkorrelationen zwischen den kognitiven Fähigkeiten erklärt. Die Spezifikation allgemeiner kognitiver Fähigkeit als Faktor 2. Ordnung impliziert, dass die allgemeine kognitive Fähigkeit die stoffgebietspezifischen Maße zusätzlich zu einer spezifischen mathematischen Fähigkeit (repräsentiert durch den Residualterm „ $M'_{HO}$ “) indirekt beeinflusst.

Gustafsson (1984) betrachtet die allgemeine kognitive Fähigkeit und die fluide Fähigkeit als äquivalent. Diese Annahme kann mit Modell 6a getestet werden, indem die Residualvarianz der fluiden Fähigkeit auf Null gesetzt wird. Dies impliziert, dass die Kovariation von den Maßen fluider Fähigkeit nur durch die allgemeine kognitive Fähigkeit erklärt wird.

*Modell 6b: „Higher-Order“-Faktorenmodell-verbal.* Als einziger Unterschied zu Modell 6a, wird in diesem Modell angenommen, dass die stoffgebietspezifischen Fähigkeiten auch von einer generellen verbalen Fähigkeit (V) abhängig sind.

*Modell 7a: Nested-Faktormodell.* Dieses Modell folgt dem Strukturmodell von Vernon, dem BIS-Modell, Carrolls Drei-Stratum-Theorie und den späteren Arbeiten von Gustafsson (1994; Gustafsson & Balke, 1993). Interindividuelle Unterschiede bei Maßen mathematischer Schülerleistung werden (multipel bedingt) direkt durch drei Fähigkeiten erklärt: allgemeine kognitive Fähigkeit (g), eine mathematikspezifische Fähigkeit (M') und stoffgebietspezifische Fähigkeiten (Residualterme).

*Modell 7b: Nested-Faktormodell-verbal.* Alle Annahmen in diesem Modell sind analog zum Modell 7a. Einziger Unterschied ist, dass in diesem Modell geprüft wird, ob interindividuelle Unterschiede bei den Maßen mathematischer Schülerleistung zusätzlich durch eine spezifische verbale Fähigkeit (V') erklärt werden können.

Abschließend sollen die Gemeinsamkeit aller Modelle und die zentralen Unterschiede zusammenfasst werden. Gemein ist allen Modellen die Annahme, dass stoffgebietspezifische Fähigkeiten einen Teil der interindividuellen Unterschiede in den Modellen erklären. Ein wesentlicher Unterschied besteht darin, ob darüber hinaus nur die allgemeine kognitive Fähigkeit (Modell 4), nur eine generelle mathematische Fähigkeit (Modell 5a) oder die allgemeine kognitive Fähigkeit *und* eine spezifische mathematische Fähigkeit ( $M'_{HO}$  in Modell 6a, bzw. M' in Modell 7a) die Maße mathematischer Schülerleistung beeinflussen. Ein weiterer Unterschied ist, ob eine verbale Fähigkeit darüber hinaus keinen Einfluss (Modell 4, Modell 5a, Modell 6a, Modell 7a) oder bedeutsamen Einfluss (Modell 5b, Modell 6b, Modell 7b) auf die interindividuellen Unterschiede bei Maßen mathematischer Schülerleistung hat.

## 7.2.2 Methode

*Stichprobe.* Als Stichprobe wurde die Analysestichprobe mit Daten von 29.386 Schülern verwendet (für die Beschreibung siehe Abschnitt 7.1.2).

*Messinstrumente und manifeste Variablen.* Da Modell 3 den besten Modell-Fit hatte, wurden als manifeste Maße mathematischer Schülerleistung Indikatoren stoffgebietsspezifischer Fähigkeiten (Ari = Arithmetik, Alg = Algebra, Geo = Geometrie, Sto = Stochastik) verwendet. Für jeden Schüler wurden als beste Punktschätzer (vgl. Rost, 2004a, S. 316) seiner stoffgebietsspezifischen Fähigkeit jeweils *Weighted-Likelihood-Estimates* (WLE, Warm, 1989) mit dem Programm ConQuest (Wu u.a., 1998) geschätzt.<sup>20</sup> Die WLE-Scores wurden nachträglich für unterschiedliche Schwierigkeiten der Stoffgebiete korrigiert, die in Abhängigkeit der bearbeiteten internationalen Testhefte entstanden (siehe Anhang; vgl. Adams & Wu, 2002).

Als Indikatoren für die fluide Fähigkeit werden im Allgemeinen Aufgaben betrachtet, die schlussfolgerndes Denken erfordern (Carroll, 1993; Horn & Noll, 1997). In der vorliegenden Arbeit wurden Figurenanalogien (FA) und Wortanalogien (WA) aus dem *Kognitiven Fähigkeitstest* (KFT, Heller & Perleth, 2000) verwendet, die bei PISA 2000 eingesetzt wurden. Bei Items aus der Skala *Figurenanalogien* werden Figurenanordnungen vorgegeben. Dabei verhält sich die erste Figur zu einer zweiten wie die dritte zu einer von fünf Alternativen. Die richtige Alternative ist herauszufinden. Items aus der Skala *Wortanalogien* erfordern eine Ergänzung von Wortanalogien der Art  $A : B = C : ?$ . Aus fünf Alternativen ist die korrekte Antwort auszuwählen. Bei beiden Aufgaben wurden richtige Antworten mit 1, falsche Antworten und fehlende Antworten mit 0 gewertet. Ausgehend von den Antworten bei den Einzelitems wurden mit dem Programm Parscale (Muraki & Bock, 1996) jeweils für beide Aufgaben WLE-Scores auf Grundlage eines zwei-parametrischen Birnbaummodells geschätzt<sup>21,22</sup> (siehe Tab. 10 für einen Vergleich der Modell-Fit-Indizes zwischen einem eindimensionalen Raschmodell und einem eindimensionalen, zweiparametrischen Birnbaummodell). Diese WLE-Scores gingen als manifeste Variablen in die Analysen ein.

Maße der verbalen Fähigkeiten waren in Anlehnung an die bei Horn und Noll (1997) und Carroll (1993) aufgelisteten Markiervariablen die Subskalen der Leseleistung, die im Rahmen von PISA unterschieden wurden. Aufgaben zur Messung der *Informationsentnahme* (Info) erfordern das Ermitteln von Informationen aus kontinuierlichen oder diskontinuierlichen (das sind z.B. Tabellen) Texten. Bei Aufgaben zum *textbezogenen Interpretieren* (Int) müssen Schüler Schlussfolgerungen aus einem oder mehreren Teilen des Textes ziehen und Textteile dahingehend beurteilen, ob sie mit bestimmten Interpretationen vereinbar sind. Wenn Aufgaben die Fähigkeit zum *Reflektieren und Bewerten* (Ref) messen, erfordert dies von den Schülern, den Aufgabentext mit eigenen Erfahrungen und ihrem Vorwissen und Ideen in Beziehung zu setzen (Artelt, Stanat, Schneider & Schiefele, 2001). Als manifeste Variablen der drei Subskalen wurden die bei PISA

---

<sup>20</sup> WLE-Parameter, die nur auf Basis von Mathematikitems mit Infit-Werten zwischen 0,8 und 1,2 geschätzt wurden (vgl. Abschnitt 7.1.3), korrelierten zwischen 0,991 und 1,0 mit WLE-Parametern, die auf Grundlage aller Mathematikitems geschätzt wurden. Eine Verwendung aller Mathematikitems „schadete“ also nicht (vgl. Köller, 1998, S. 75).

<sup>21</sup> Eines von 20 Items der Skala WA wurde bei der Berechnung der WLE-Scores ausgeschlossen, da es eine Trennschärfe von nahezu Null hatte.

<sup>22</sup> Der Versuch scheiterte, die Daten des PISA-Mathematiktests mit Parscale zu analysieren: Für einige Modellparameter resultierten unrealistische Werte (z.B. Diskriminationsparameter von Null, trotz Trennschärfeparameter der klassischen Testtheorie, die ungleich Null waren), extreme Itemschwierigkeiten oder sehr große Standardfehler der Modellparameter. Es ist anzunehmen, dass die Anzahl zufälliger fehlender Werte, die durch das Matrixdesign resultierten, zu groß war, um mit Parscale verlässliche Modellparameter zu schätzen.

2000 geschätzten WLE-Scores verwendet, die sich auf der PISA-Metrik befanden ( $M = 500$ ,  $SD = 100$ ). Um deren Varianzen an die Varianzen der Indikatoren für stoffgebietspezifische Fähigkeiten und für die fluide Fähigkeit anzugleichen, wurden die WLE-Scores der Leseleistung wieder auf die Logit-Metrik retransformiert. Damit wurde vorgebeugt, dass zu große Varianzunterschiede zu verzerrten Schätzungen der Modellparameter führten.

Tabelle 10: Fluide Fähigkeit: Vergleich der Modll-Fit-Indizes für ein eindimensionales Rasch-Modell (1p) und ein eindimensionales, zweiparametrisches Birnbaum-Modell (2p)

Skala (Parametrisierung)	$\chi^2$	df	AIC	BIC	CAIC
Figurenanalogien (1p)	17.460	637	17.512	17.576	17.602
Figurenanalogien (2p)	12.210	1.522	<b>12.310</b>	<b>12.434</b>	<b>12.484</b>
Wortanalogien (1p)	8.821	354	8.861	8.910	8.930
Wortanalogien (2p)	7.603	1.184	<b>7.679</b>	<b>7.773</b>	<b>7.811</b>

df = Freiheitsgrade. Die kleinsten Werte von AIC = *Aikaikes Information Criterion*, BIC = *Bayes Information Criterion* und CAIC = *Consistent AIC* sind fett gedruckt. Für ein dreiparametrisches Birnbaummodell wurden keine Fit-Indizes ausgegeben, was die Robustheit der geschätzten Modellparameter infrage stellte. Daher wurde es nicht weiter berücksichtigt.

*Reliabilitätsberechnung für die WLE-Scores.* Reliabilitäten für die WLE-Scores wurden geschätzt, indem Maße der „wahren“ Varianz durch die gesamte Varianz der WLE-Scores geteilt wurden. Bei den stoffgebietspezifischen Fähigkeiten wurde die „wahre“ Varianz direkt als die Varianz der stoffgebietspezifischen Faktoren in Modell 3 bestimmt (siehe Tab. 8). Zur Schätzung der Reliabilität wurde diese latente Varianz durch die Varianz der WLE-Scores geteilt (siehe Rost, 2004a, Formel 6, S. 381). Diese Methode war bei den anderen WLE-Scores nicht anwendbar, da keine Schätzungen der latenten Varianzen für die verwendete Analysestichprobe vorlagen (in Parscale wurde diese für die fluiden Fähigkeitsmaße nicht angegeben; die latenten Varianzen der Lesesubskalen von PISA 2000 lagen zwar vor, bezogen sich aber nicht auf die Analysestichprobe dieser Arbeit). Daher wurde die „wahre“ Varianz indirekt bestimmt (Rost, 2004a, Formel 5, S. 380). Als „wahre“ Varianz wurde die Differenz zwischen der Gesamtvarianz der WLE-Scores und der mittleren Varianz der Standardschätzfehler der WLE-Scores (als Maß der Fehlervarianz) berechnet. Teilt man diese Differenz durch die Gesamtvarianz der WLE-Scores, erhält man eine Schätzung der Reliabilität.

*Konfirmatorische Faktorenanalyse.* Alle Modelle wurden mit konfirmatorischen Faktorenanalysen analysiert. Zur Identifikation wurde bei allen Faktoren die unstandardisierte Ladung eines Indikators auf 1,0 fixiert. Zusätzlich wurde in den Modellen 5a/b und 6a/b zur Identifikation von Gf jeweils die Ladung von FA auf Gf und von WA auf Gf auf 1,0 fixiert. Alle anderen Modellparameter wurden frei mit dem Programm Mplus 3.01 (Muthén & Muthén, 1998–2004b) geschätzt. Bei der Schätzung der Modellparameter und der korrespondierenden Standardfehler mussten zwei Besonderheiten der Daten beachtet werden:

- Erstens die genestete Datenstruktur, die aus der mehrstufigen Stichprobenziehung resultierte (vgl. Abschnitt 7.1.2).

- Zweitens hatten einige Schüler fehlende Werte bei einer oder mehreren manifesten Variablen (vgl. Tab. 11 und Anhang): Maximal fehlten von 146 Schülern die WLE-Scores zur Messung fluider Fähigkeit; insgesamt hatten 29.231 von 29.386 Schüler gültige Werte bei allen manifesten Variablen.

Aus diesen beiden Gründen wurden die individuellen Schülerrohdaten mit dem Modul „complex“ und dem Schätzalgorithmus MLR analysiert, da damit gleichzeitig die genestete Datenstruktur und die fehlenden Werte berücksichtigt wurden.<sup>23</sup>

Zur Evaluation des Modell-Fits wurden mehrere Kriterien herangezogen:

- (a) Beim  $\chi^2$ -*Goodness-of-Fit*-Test wird die Null-Hypothese getestet, dass das spezifizierte Modell das „wahre“ Populationsmodell ist. Ein nicht signifikanter  $\chi^2$ -Test bedeutet, dass – unter der Annahme, dass das spezifizierte Modell das „wahre“ Populationsmodell ist – die beobachtete Varianz-Kovarianzmatrix nicht signifikant von der modell-implizierten Varianz-Kovarianzmatrix abweicht. Jedoch ist – wie auch bei den IRT-Analysen – einer der größten Nachteile dieses Tests, dass selbst marginale Abweichungen zwischen empirischer und modell-implizierter Varianz-Kovarianzmatrix signifikant werden, wenn große Stichproben verwendet werden (z.B. Kaplan, 1990).
- (b) Aufgrund der Sensitivität des  $\chi^2$ -*Goodness-of-Fit*-Tests wurden zusätzlich deskriptive Maße zur Evaluierung der Modelle herangezogen. Der *Comparative Fit Index* (CFI, Bentler, 1990) vergleicht den inkrementellen Zuwachs an Modell-Fit im Vergleich zu einem Modell, in dem die Kovarianzen der manifesten Variablen auf Null fixiert und nur die Varianzen frei geschätzt werden. Ein Vertreter absoluter Fit-Indizes ist der *Standardized Root Mean Square of Residuals* (SRMR) Index (z.B. Bentler, 1995). Der SRMR gibt an, wie genau das spezifizierte Modell die Daten reproduziert. Ein Problem bei alleiniger Verwendung des SRMR ist, dass das komplexeste Modell immer den besten Fit haben wird, wenn die Anzahl frei geschätzter Parameter und damit die Modellkomplexität zunimmt. Die Sparsamkeit eines Modells (*parsimony*) wird somit beim SRMR nicht berücksichtigt. Ein Fit-Index, der den absoluten Modell-Fit, die *parsimony* des Modells und die Stichprobengröße gleichzeitig berücksichtigt, ist der *Root Mean Square Error of Approximation* (RMSEA, z.B. Steiger, 1990).

Für die deskriptiven Fit-Indizes werden eine Reihe von so genannten *Cut-off*-Werten vorgeschlagen, wann ein Modell als gute Approximation zu betrachten ist. Hu und Bentler (1998) geben als *Cut-off*-Werte für den CFI .95, für den SRMR .08 und für den RMSEA .06 an. In einer vielzitierten Arbeit bewerten Browne und Cudeck (1993, S. 144) Werte des RMSEA kleiner .05 als „*a close fit of the model in relation to the degrees of freedom*“; Werte kleiner .08 als „*reasonable error of approximation*“ und sagen, dass sie kein Modell verwenden würden „*with a RMSEA greater than 0.1*“.

- (c) Die alleinige Verwendung von *Cut-off*-Werten wird aber auch sehr kritisch betrachtet (z.B. Marsh, Hau & Wen, 2004). McDonald (1999) und Rindskopf (2000) empfehlen zur Evaluie-

---

<sup>23</sup> Der Schätzalgorithmus MLR gehört zur Familie der *Full Information Maximum Likelihood Verfahren* (FIML). Diese können Populationsstatistiken (z.B. Mittelwerte, Varianzen und Kovarianzen) auf Grundlage unvollständiger Daten schätzen. Bei Verwendung der FIML-Verfahren wird angenommen, dass das Fehlen eines Werts eine Funktion der anderen analysierten Variablen ist (*Missing at Random* [MAR], Rubin, 1987). Die zentrale Idee der FIML-Verfahren bei der Parameterschätzung ist, dass (unter der Annahme von MAR) auf Grundlage der tatsächlich beobachteten Daten die Likelihood für die unbekanntenen Populationsparameter maximiert wird (Schafer & Graham, 2002).

rung von Modellen, rivalisierende Modelle miteinander zu vergleichen. McDonald schlägt vor, das Modell zu favorisieren, das den besten Modell-Fit aufweist und gleichzeitig aus theoretischer Sicht interpretiert werden kann. Ein Indikator der theoretischen Interpretierbarkeit ist die Substanz der Faktorladungen, die nach Carroll (1993) und McDonald (1999) mit Werten der standardisierten Faktorladungen größer oder gleich .30 gegeben ist.

Bei allen inferenzstatistischen Analysen wurde ein Signifikanzniveau von  $p < .05$  gesetzt.

### 7.2.3 Ergebnisse

*Deskriptive Statistiken.* Die deskriptiven Statistiken der verwendeten manifesten Variablen sind in Tabelle 11 eingetragen. Daraus ist ersichtlich, dass alle manifesten Variablen positiv interkorrelierten, was die Annahme einer allgemeinen kognitiven Fähigkeit stützte. Interessant war dabei, dass der WLE-Score für Stochastik deutlich geringer mit den anderen kognitiven Fähigkeitsmaßen korrelierte. Dies ist kein Reliabilitätsproblem des Stochastik-WLE-Scores, da deren Reliabilität in ähnlicher Höhe lag wie die von Geometrie oder Arithmetik.<sup>24</sup> Die geringeren Interkorrelationen sprachen vielmehr dafür, dass – im Vergleich zu den anderen Stoffgebieten – die Leistung bei Stochastikaufgaben deutlich stärker von der Fähigkeit zur Anwendung stochastikspezifischen Wissens war.

Tabelle 11: Deskriptive Statistiken der WLE-Scores zur Messung kognitiver Fähigkeiten

	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. Algebra	<i>0,84</i>								
2. Arithmetik	0,58	<i>0,71</i>							
3. Geometrie	0,52	0,53	<i>0,61</i>						
4. Stochastik	0,40	0,39	0,33	<i>0,66</i>					
5. Figurenanalogien	0,51	0,48	0,47	0,32	<i>0,91</i>				
6. Wortanalogien	0,52	0,48	0,44	0,33	0,55	<i>0,85</i>			
7. Information	0,52	0,50	0,44	0,37	0,48	0,53	<i>0,57</i>		
8. Interpretation	0,54	0,52	0,45	0,39	0,50	0,57	0,71	<i>0,73</i>	
9. Reflektion	0,47	0,44	0,38	0,32	0,43	0,49	0,59	0,67	<i>0,53</i>
M	0,09	0,01	-0,03	0,24	0,03	0,04	-0,08	-0,06	-0,17
SD	1,41	1,36	1,41	1,18	1,08	1,27	1,09	1,05	1,20
#fehlend	3	3	3	3	146	146	1	8	8

Korrelationen, Mittelwerte (M) und Standardabweichungen (SD) beziehen sich auf die mit dem FIML-Schätzer ermittelten Stichprobenkennwerte. In der Diagonale sind die Reliabilitäten der WLE-Scores eingetragen (für ihre Berechnung siehe Abschnitt 7.2.2). #fehlend = absolute Anzahl fehlender Werte.

<sup>24</sup> Teilweise waren die Interkorrelationen der Indikatoren der verbalen Fähigkeit größer, als dies ihre Reliabilität theoretisch zuließ (Lord & Novick, 1968; Schmitt, 1996). Ein Grund hierfür war möglicherweise, dass die Reliabilitäten durch das verwendete Berechnungsverfahren unterschätzt wurden, weil die Fehlervarianzen überschätzt wurden (siehe hierzu Rost, 2004b).

*Evaluation des Modell-Fits.* Bei der Schätzung der Parameter für die Modelle 5b und 6b gab es Schwierigkeiten. Bei Modell 5b wurden keine Standardfehler berechnet, bei Modell 6b war die standardisierte Faktorladung von Gf auf g größer 1,0. Diese Schätzprobleme sind nicht identifikationsbedingt, sondern wahrscheinlich modellinhärent: Die Modelle 5b und 6b sind Modellen zur Analyse von *Multitrait-Multimethod*-Matrizen sehr ähnlich. Insbesondere bei Letzteren sind Probleme bei der Parameterschätzung hinlänglich bekannt (z.B. Marsh, 1989). Da keine gültigen Parameterschätzungen für die Modelle 5b und 6b vorlagen, wurden diese beiden Modelle nicht weiter betrachtet.

Inspizierte man die globalen Modell-Fit-Indizes (Tab. 12) der anderen Modelle, so war bei allen Modellen der  $\chi^2$ -Wert signifikant. Jedoch konnten, gemessen an den *Cut-off*-Werten, die Modelle 5a, 6a, 7a und 7b als gute bis sehr gute Approximationen der empirischen Daten betrachtet werden. Vergleich man die deskriptiven Fit-Indizes (CFI, RMSEA, SRMR) dieser Modelle miteinander, so deuteten alle Indizes darauf hin, dass Modell 7b die empirischen Relationen zwischen den manifesten Variablen am besten erklären konnte.

Wie die großen Unterschiede zwischen den  $\chi^2$ -Werten bereits augenscheinlich nahe legten, waren die Differenzen ( $\Delta\chi^2$ ), die entsprechend der Korrekturformel aus dem technischen Anhang von Mplus (Muthén & Muthén, 1998–2004a, S. 22) berechnet wurden, zwischen den genesteten Modellen 4 und 5a ( $\Delta\chi^2 = 19.597$ ,  $df = 2$ ), Modellen 4 und 6a ( $\Delta\chi^2 = 19.597$ ,  $df = 2$ ) sowie Modellen 4 und 7a ( $\Delta\chi^2 = 8.941$ ,  $df = 7$ ) statistisch signifikant. Dies bedeutete, dass die Modelle 5a, 6a und 7a die empirischen Relationen zwischen den manifesten Variablen besser erklären konnten als Modell 4, in dem nur ein Faktor für allgemeine kognitive Fähigkeit spezifiziert wurde. Auch die  $\chi^2$ -Differenz der Modelle 7a und 7b ( $\Delta\chi^2 = 232$ ,  $df = 4$ ) war statistisch signifikant. Dies stützte zunächst die Hypothese, dass eine spezifische verbale Fähigkeit zusätzlich zur allgemeinen kognitiven Fähigkeit und der mathematikspezifischen Fähigkeit interindividuelle Unterschiede bei Maßen mathematischer Schülerleistung erklären konnte.

Tabelle 12: Globale Modell-Fit-Indizes

Modell	$\chi^2$	df	SF	CFI	RMSEA	SRMR
Modell 4: Allgemeine kognitive Fähigkeit	8.406	27	0,989	0,93	0,10	0,04
Modell 5a: Standardmodell (Gf-Gc-Modell)	1.370	25	1,040	0,99	0,04	0,03
Modell 6a: <i>Higher-Order</i> -Faktorenmodell	1.370	25	1,040	0,99	0,04	0,03
Modell 7a: Nested-Faktormodell	672	20	1,037	1,00	0,03	0,01
Modell 7b: Nested-Faktormodell-verbal	430	16	1,015	1,00	0,03	0,01

SF = Skalierungsfaktor, CFI = *Comparative Fit Index*, RMSEA = *Root Mean Square Error of Approximation*, SRMR = *Standardized Root Mean Square of Residuals*. Die Modelle 5b und 6b waren empirisch unteridentifiziert. Daher sind für diese Modelle keine Fitstatistiken eingetragen.

*Modellparameter.* Zog man allerdings noch als weiteres Kriterium der Modellevaluation die Substanz der standardisierten Faktorladungen hinzu (Tab. 13), dann wurde offensichtlich, dass die Ladungen der stoffgebietspezifischen WLE-Scores auf V' in Modell 7b nicht substantiell waren. Drei der vier Faktorladungen waren zwar statistisch signifikant von Null verschieden, doch die Werte lagen nur zwischen .02 und .13. Der bessere Fit von Modell 7b ist also in erster Linie das



Ergebnis einer Überparametrisierung (*overfitting*). Aus diesem Grund wurde Modell 7b nicht weiter betrachtet.

Tabelle 13: Standardisierte Faktorladungen ( $\lambda$ ) der Modelle 4 bis 7b

	Modell 4		Modell 5a			Modell 6a			Modell 7a			Modell 7b		
	$\lambda_g$	$\lambda_M$	$\lambda_{Gf}$	$\lambda_V$	$\lambda_M$	$\lambda_{Gf}$	$\lambda_V$	$\lambda_{M'}$	$\lambda_g$	$\lambda_{V'}$	$\lambda_{M'}$	$\lambda_g$	$\lambda_{V'}$	
Alg	.72 <sup>a</sup>	.77 <sup>a</sup>			.77 <sup>a</sup>			.25 <sup>a</sup>	.72 <sup>a</sup>		.29 <sup>a</sup>	.70 <sup>a</sup>	.08	
Ari	.69	.75			.75			.39	.68		.41	.65	.09	
Geo	.62	.68			.68			.30	.62		.31	.61	.02	
Sto	.49	.52			.52			.17	.48		.23	.44	.13	
FA	.66		.76 <sup>a</sup>		.76 <sup>a</sup>				.70			.72		
WA	.71		.70 <sup>a</sup>		.70 <sup>a</sup>				.74			.76		
Info	.78			.81 <sup>a</sup>			.81 <sup>a</sup>		.71	.35 <sup>a</sup>		.69	.40 <sup>a</sup>	
Int	.82			.88			.88		.75	.49		.73	.52	
Ref	.71			.75			.75		.64	.39		.62	.42	

Alle Faktorladungen waren (mit Ausnahme von Geo auf  $V'$  in Modell 7b) statistisch signifikant von Null verschieden. Alg = Algebra, Ari = Arithmetik, Geo = Geometrie, Sto = Stochastik, FA = Figurenanalogien, WA = Wortanalogien, Info = Informationen ermitteln, Int = textbezogenes Interpretieren, Ref = Reflektieren und Bewerten.

<sup>a</sup> Die unstandardisierte Faktorladung dieser manifesten Variable wurde im jeweiligen Modell auf 1,0 fixiert.

In den Modellen 4, 5a und 6a waren alle Faktorladungen substanziell und signifikant von Null verschieden. Bei Modell 7a, das gemessen an den deskriptiven Modell-Fit-Indizes den zweitbesten Fit hatte, waren alle Ladungen auf  $g$  und  $V'$  größer .30. Die mathematikspezifische Fähigkeit hatte einen schwachen bis moderaten Einfluss auf die stoffgebietsspezifischen Fähigkeiten: Zwei von vier Ladungen auf  $M'$  waren gleich oder größer .30, die beiden anderen waren zumindest statistisch signifikant von Null verschieden. Modell 7a wurde daher sowohl aufgrund des globalen Modell-Fits als auch der besseren theoretischen Interpretierbarkeit favorisiert.

In Modell 5a betrug die latente Korrelation zwischen  $Gf$  und  $M$  .89, zwischen  $Gf$  und  $V$  .83 sowie zwischen  $M$  und  $V$  .80. Diese hohen Korrelationen stützten die Annahme, dass die Interkorrelationen auf die investierte fluide Fähigkeit attribuiert werden können. Allerdings stützten diese Korrelationen auch die Interpretation, dass sie den Einfluss der allgemeinen kognitiven Fähigkeit repräsentierten.

Letzteres wurde auch direkt durch die hohen standardisierten Faktorladungen in Modell 6a bestätigt:  $Gf$  lud auf  $g$  mit .96,  $M$  lud auf  $g$  mit .92 und  $V$  lud auf  $g$  mit .86.

Um zu prüfen, ob  $Gf$  und  $g$  in Modell 6a äquivalent waren, wurde die Residualvarianz von  $Gf$  auf Null fixiert. Der korrespondierende  $\chi^2$ -Differenzentest war signifikant ( $\Delta\chi^2 = 95$ ,  $\Delta df = 1$ ). Das heißt, aus inferenzstatistischer Sicht konnte die Äquivalenzannahme nicht gestützt werden. Betrachtete man aber die Höhe der standardisierten Faktorladung von .96 und den Varianzanteil von 92,3 Prozent, den  $g$  an  $Gf$  erklärte, so konnte man zumindest aus deskriptiver Sicht nahezu von der Identität von  $Gf$  und  $g$  sprechen.

*Varianzanteile.* Abschließend wurde noch der Frage nachgegangen, wie stark der Einfluss der kognitiven Fähigkeiten auf die stoffgebietspezifischen WLE-Scores war. Die Faktorladungen in Modell 7a konnten so interpretiert werden, dass interindividuelle Unterschiede in der allgemeinen kognitiven Fähigkeit, in der mathematikspezifischen Fähigkeit und in den stoffgebietspezifischen Fähigkeiten die interindividuellen Unterschiede bei den stoffgebietspezifischen WLE-Scores erklärten. Der Einfluss von  $g$  und  $M'$  auf die Leistung bei den stoffgebietspezifischen WLE-Scores konnte direkt über die Faktorladungen ermittelt werden: Je höher die Faktorladung, umso stärker war der Einfluss.

Welchen Anteil der Varianz erklärten die stoffgebietspezifischen Fähigkeiten? Dieser Varianzanteil konnte nicht direkt anhand der Modellparameter berechnet werden, da sich die Residualvarianzen der manifesten WLE-Scores zusammensetzten aus stoffgebietspezifischen Anteilen und durch Anteile, die durch Schätzfehler bedingt waren (vgl. Bollen, 1989). Jedoch konnte der Einfluss stoffgebietspezifischer Fähigkeiten indirekt über die Reliabilität der WLE-Scores *abgeschätzt* werden. Zog man vom reliablen Varianzanteil eines WLE-Scores den Anteil ab, der durch die allgemeine kognitive Fähigkeit und die mathematikspezifische Fähigkeit erklärt wurde (dies ist die Summe der beiden quadrierten Faktorladungen, z.B. Bollen, 1989), erhielt man eine Schätzung des Varianzanteils, der auf die jeweilige stoffgebietspezifische Fähigkeit zurückging.

In Abbildung 17 sind die prozentualen Varianzanteile der jeweiligen Fähigkeit an der Gesamtvarianz eines WLE-Scores dargestellt. Es ist ersichtlich, dass die erklärten Varianzanteile stark abhängig von den jeweiligen Stoffgebieten waren. So erklärte die allgemeine kognitive Fähigkeit (hellgrauer Balken) 51 Prozent der Varianz des Algebra-WLE-Scores, aber nur 23 Prozent der Varianz des Stochastik-WLE-Scores. Die mathematikspezifische Fähigkeit (weißer Balken) erklärte 15 Prozent der Varianz des Arithmetik-WLE-Scores, aber nur 3 Prozent des Stochastik-WLE-Scores. Hingegen erklärten stoffgebietspezifische Fähigkeiten (dunkelgrauer Balken) 40 Prozent des Stochastik-WLE-Scores, aber nur 10 Prozent des Arithmetik-WLE-Scores.

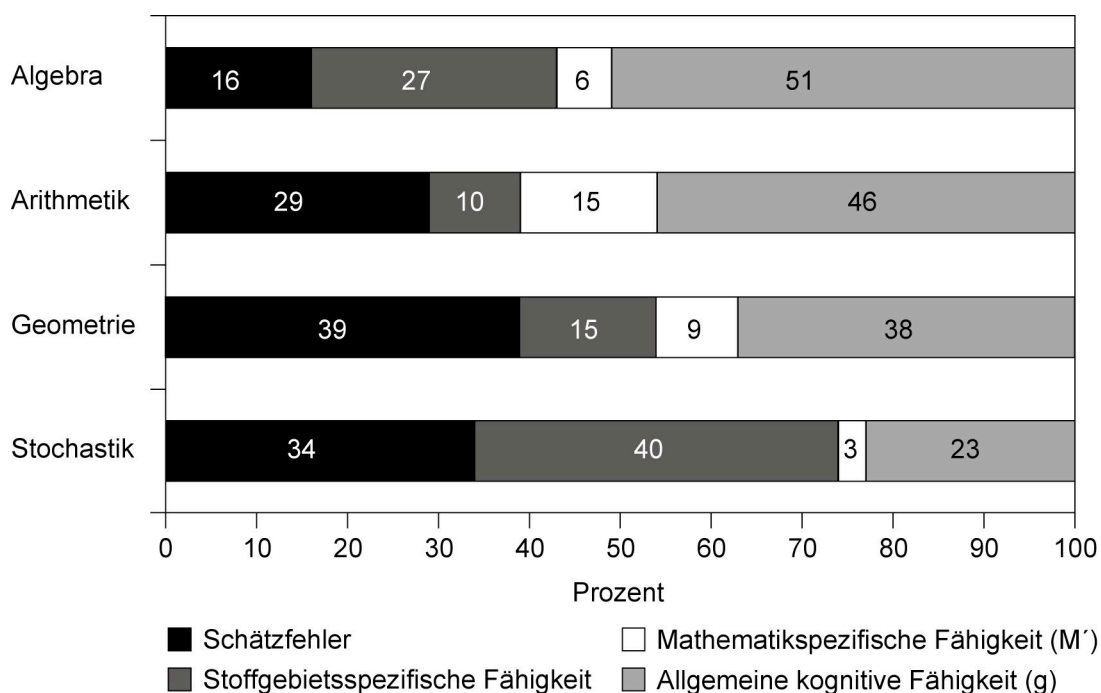


Abbildung 17: Modell 7a – Varianzzusammensetzung der stoffgebietspezifischen WLE-Scores

## 7.2.4 Diskussion

Die beiden wichtigsten Ergebnisse der vorangegangenen Analysen können wie folgt zusammengefasst werden:

- Erstens wurden interindividuelle Unterschiede bei Maßen mathematischer Schülerleistung erklärt durch das (additiv-kompensatorische) Zusammenwirken von interindividuellen Unterschieden in den stoffgebietspezifischen Fähigkeiten, in der mathematikspezifischen Fähigkeit und in der allgemeinen kognitiven Fähigkeit.
- Zweitens war die verbale Fähigkeit kein bedeutsamer Erklärungsfaktor für interindividuelle Unterschiede bei Maßen mathematischer Schülerleistung.

Nachfolgend werden zunächst Überlegungen zum ersten Ergebnis dargestellt, das aus theoretischer Sicht in mehrerlei Hinsicht interessant ist. Das zweite Ergebnis wird im Rahmen der Gesamtdiskussion in Abschnitt 7.3 behandelt.

Im Rahmen der Modelltests wurden aus Strukturtheorien kognitiver Fähigkeiten Annahmen abgeleitet und überprüft, welche kognitiven Fähigkeiten interindividuelle Unterschiede bei Maßen mathematischer Schülerleistung erklären können. Die alleinige Annahme einer allgemeinen kognitiven Fähigkeit (Modell 4) konnte empirisch nicht bestätigt werden. Das Standardmodell (Modell 5a), das die Theorie fluider und kristalliner Fähigkeiten (Horn & Noll, 1997) abbildete, wies eine gute Modellpassung auf. Es erklärte aber die empirischen Relationen – auch unter Berücksichtigung der Sparsamkeit des Modells (z.B. gemessen am RMSEA) – schlechter als das rivalisierende Nested-Faktormodell (Modell 7a). Dies sprach gegen die Annahme in der Theorie fluider und kristalliner Fähigkeiten, dass keine allgemeine kognitive Fähigkeit existiert (siehe auch Carroll, 2003).

Im Nested-Faktormodell wurde die multiple Bedingtheit der stoffgebietspezifischen Maße mathematischer Schülerleistung aus den Annahmen des Modells von Vernon (1964), dem BIS-Modell (Jäger u.a., 1997) und Carrolls (1993) Drei-Stratum-Theorie abgeleitet. Die empirischen Ergebnisse stützten diese Modellannahmen: Interindividuelle Unterschiede in stoffgebietspezifischen Maßen mathematischer Schülerleistung werden durch das Zusammenwirken von drei verschiedenen kognitiven Fähigkeiten (allgemeine kognitive Fähigkeit, mathematikspezifische Fähigkeit und stoffgebietspezifische Fähigkeiten) erklärt. Somit werden auch zwei der Erklärungen für die latenten Interkorrelationen zwischen den stoffgebietspezifischen Fähigkeiten empirisch gestützt (vgl. Abschnitt 7.1.4). Die latenten Korrelationen spiegeln sowohl den Einfluss der allgemeinen kognitiven Fähigkeit *als auch* den Einfluss der mathematikspezifischen Fähigkeit wider.

Weiterhin unterfüttern die Ergebnisse (vgl. Modell 6a) die Annahme von Gustafsson (1984), dass die fluide Fähigkeit und die allgemeine kognitive Fähigkeit weitestgehend äquivalent sind. Die allgemeine kognitive Fähigkeit ( $g_{HO}$ ) erklärte nahezu vollständig die Varianz der fluiden Fähigkeit. Hinsichtlich der 7,7 Prozent nicht erklärter Varianz ist auch zu bedenken, dass im Gegensatz zu der Studie von Gustafsson in PISA 2000 (unter anderem aufgrund der begrenzten Testzeit von 30 Minuten) nur zwei Markieraufgaben der fluiden Fähigkeit verwendet wurden und somit das Konstrukt „fluide Fähigkeit“ nicht optimal repräsentiert war (z.B. wurde keine Markieraufgabe zum schlussfolgernden Denken mit numerischem Inhalt in PISA eingesetzt).

Diese Überlegung ist umso wichtiger vor dem Hintergrund der oftmaligen Trennung der Konstrukte und Theorien aus der Strukturforschung kognitiver Fähigkeit einerseits und Schülerleistungen andererseits (vgl. Abschnitt 4.1). Nur wenn alle Konstrukte bestmöglich operationalisiert werden, kann man ein unverzerrtes Bild über deren Zusammenwirken zeichnen: Das Nested-Faktormodell (Modell 7a) ist hierfür ein gutes Beispiel und belegt die Tragfähigkeit eines „coordinate measurement of developed cognitive abilities along with subject-matter achievement“, wie es Messick (1984, S. 226) anregte. Mithilfe des Nested-Faktormodells konnte ein differenzierter Einblick gewonnen werden, wie sich die Varianz bei Maßen mathematischer Schülerleistung zusammensetzt. Das wichtigste Ergebnis dabei war (vgl. Abb. 17), dass – entgegen bisheriger Annahmen (Klieme, Neubrand u.a., 2001, S. 156) – die allgemeine kognitive Fähigkeit, die mathematischspezifische Fähigkeit und stoffgebietsspezifische Fähigkeiten nicht gleichermaßen Einfluss auf die Varianzzusammensetzung von Maßen mathematischer Schülerleistung nahmen.

In diesem Zusammenhang ist eine wichtige Frage, wie die deutlich unterschiedlichen Zusammenhänge zwischen allgemeiner kognitiver Fähigkeit sowie der mathematikspezifischen Fähigkeit und den Maßen mathematischer Schülerleistung erklärt werden können (vgl. Abb. 17). Hierbei sollten drei miteinander verquickte Aspekte beachtet werden: Selbstselektionsprozesse in höhere Schulformen, schulformspezifische Unterschiede des Mathematikunterrichts und der Prozessablauf beim mathematischen Problemlösen.

- (a) Schüler mit stärker ausgeprägter allgemeiner kognitiver Fähigkeit besuchen meist höhere Schulformen. Hierbei ist bekannt, dass an höheren Schulformen mathematische Fähigkeiten besser gefördert werden als an niedrigeren Schulformen (vgl. Abschnitt 5.4). Der enge Zusammenhang zwischen der allgemeinen kognitiven Fähigkeit und Maßen mathematischer Schülerleistung ist daher auch durch diesen Selektionseffekt bedingt.
- (b) Weiterhin muss man auch Spezifika des deutschen Mathematikunterrichts und die Tatsache beachten, dass differenzielle Unterschiede der Lernumwelten Einfluss auf die Faktorstruktur mathematischer Schülerleistung nehmen können (vgl. Abschnitt 4.3.1). Beobachtungen des deutschen Mathematikunterrichts zeigten, dass schulformübergreifend an Deutschlands Schulen generell verstärkt technische Aufgaben und mit deutlich geringeren Anteilen problemorientierte rechnerische oder begriffliche Modellierungsaufgaben eingesetzt werden (Klieme, Schümer u.a., 2001; Kunter, 2005). Diese Tendenz zur Einheitlichkeit des deutschen Mathematikunterrichts impliziert, dass sich die Lernumwelten hinsichtlich der Förderung einer problemorientierten mathematischen Fähigkeit (wie sie z.B. in Form der mathematikspezifischen Fähigkeit konzeptualisiert werden kann, vgl. Abschnitt 7.3) nicht so stark voneinander unterscheiden. Aufgrund dieser Varianzrestriktion ist es nicht so überraschend, dass nur relativ geringe Anteile der interindividuellen Unterschiede durch die mathematikspezifische Fähigkeit erklärt werden.
- (c) Der verstärkte Einsatz technischer Aufgaben im deutschen Mathematikunterricht bedeutet auch, dass Schüler in Deutschland tendenziell weniger mit Aufgaben zum rechnerischen und begrifflichen Modellieren vertraut sind (siehe auch Klieme & Baumert, 2001). Diese Überlegung ist wichtig vor dem Hintergrund des Phasenmodells des Fertigkeitserwerbs, da sich Schüler aller Schulformen hinsichtlich dieser Aufgaben tendenziell ganz zu Beginn der kognitiven Phase sensu Ackerman (1989; siehe auch Abschnitt 5.2.1), bzw. im Problemlöseprozess sensu Anderson (1993) befinden. Unter der Annahme, dass sich Schüler zu Beginn der kognitiven Phase befinden, erklären interindividuelle Unterschiede in der allgemeinen kognitiven

Fähigkeit einen Großteil der interindividuellen Unterschiede zumindest bei den rechnerischen und begrifflichen Modellierungsaufgaben aus dem PISA-Test.

Unabhängig vom genauen Wirkmechanismus kann anhand der Befunde zur ersten Forschungsfrage ein differenzierteres Bild gezeichnet werden, welche Rolle mathematikspezifische und (nicht genuin mathematische) kognitive Fähigkeiten für die Lösung von Mathematikaufgaben aus dem jeweiligen Stoffgebiet spielen. Eine Interpretation der Varianzanteile der jeweiligen Fähigkeiten ergibt, dass mit zunehmendem Varianzanteil die Relevanz der jeweiligen Fähigkeit am Zustandekommen der stoffgebietsspezifischen Leistung (in der untersuchten Stichprobe) zunimmt. Für ein besseres Verständnis dieses Ergebnisses ist es hilfreich, diese Fähigkeiten an Konstrukte der Informationsverarbeitung und der Problemlöseforschung anzubinden. Dies wird im Rahmen der Gesamtdiskussion der ersten Forschungsfrage versucht. Vor diesem Hintergrund wird auch das zweite Teilergebnis „verbale Fähigkeit hat keinen Einfluss auf die Leistung bei Maßen mathematischer Schülerleistung“ erklärt.

### **7.3 Zusammenfassende Diskussion zur Struktur mathematischer Schülerleistung**

Angesichts des Hauptziels der vorliegenden Arbeit, mathematische Schülerleistung als psychologisches Konstrukt zu spezifizieren und es aus dieser Perspektive besser zu verstehen, ist mit den Ergebnissen zur ersten Forschungsfrage ein wichtiger Schritt getan: Interindividuelle Unterschiede in Maßen mathematischer Schülerleistung spiegeln bei einer Konzeptionalisierung in Form des Nested-Faktormodells interindividuelle Unterschiede in der allgemeinen kognitiven Fähigkeit, der mathematikspezifischen Fähigkeit und in den stoffgebietsspezifischen Fähigkeiten wider.

Um mathematische Schülerleistung noch fundierter im Raum psychologischer Konstrukte zu verorten, ist es interessant zu fragen, mit welchen Konstrukten aus der Forschung zur Informationsverarbeitung und der Problemlöseforschung die mathematikspezifische Fähigkeit und die allgemeine kognitive Fähigkeit korrespondiert. Zunächst wird auf die mathematikspezifische Fähigkeit eingegangen. Diese kann, wie in Abschnitt 7.1.4 ausgeführt wurde, als die stoffgebietsunabhängige Fähigkeit zum mathematischen Problemlösen interpretiert werden. Diese Interpretation wird nachfolgend durch Verweise auf die Literatur gestützt und weiter elaboriert. Mit dieser theoretischen Explikation wird auch der methodischen Kritik am Nested-Faktormodell begegnet, dass die spezifischen Faktoren theoretisch erklärt werden sollen (Schulze, 2005, S. 253).

*Interpretation der mathematikspezifischen Fähigkeit.* Geary (1994, S. 146) zieht zusätzlich zu stoffgebietsspezifischen Fähigkeiten zwei generellere Fähigkeiten zur Erklärung interindividueller Unterschiede bei Maßen mathematischer Schülerleistung heran. Dies sind die Fähigkeit zum Aufbau einer mathematischen Problemrepräsentation sensu Mayer (1985) und die Arbeitsgedächtniskapazität als die Fähigkeit zum simultanen Speichern und Verarbeiten von Informationen.

Der Aufbau einer mathematischen Problemrepräsentation erfordert nach Mayer (1985), den Problemtext und gegebenenfalls vorhandene Abbildungen und Grafiken in eine mentale Repräsentation zu übersetzen, problemrelevante Informationen der mentalen Repräsentation zu entnehmen und mithilfe des schematischen Mathematikwissens in eine kohärente mentale Struktur (mathematisches Problemmodell) zu integrieren. Hierzu ist mathematisches Wissen (schematisches, strategisches und prozedurales Wissen) notwendig, wovon vor allem das schematische und prozedurale Wissen vornehmlich stoffgebietsspezifisch sind. Unter dieser Annahme repräsentieren die Residualterme der WLE-Scores (zusätzlich zur Schätzfehlervarianz) interindividuelle

Unterschiede in der Verfügbarkeit und der Fähigkeit zur Anwendung des stoffgebietspezifischen Wissens.

Im Gegensatz hierzu sind die kognitiven Prozesse zum Aufbau eines (mental repräsentierten) mathematischen Problemmodells wie auch ein bedeutsamer Teil der Planungs- und Überwachungsprozesse nicht spezifisch für ein mathematisches Stoffgebiet. Schüler, die die Teilprozesse besser ausführen können als andere Schüler, sollten damit (unabhängig von der allgemeinen kognitiven Fähigkeit oder stoffgebietspezifischen Fähigkeiten) bessere Leistungen bei *allen* problemorientierten Maßen mathematischer Schülerleistung erzielen. Folgt man dieser Argumentation, dann repräsentiert die Varianz der mathematikspezifischen Fähigkeit ein Varianzamalgam, das sich aus den interindividuellen Unterschieden bei den zu Grunde liegenden Teilprozessen des mathematischen Problemlösens zusammensetzt.

*Interpretation der allgemeinen kognitiven Fähigkeit.* Als weitere Quelle interindividueller Unterschiede nennt Geary (1994) die Arbeitsgedächtniskapazität. Von der Arbeitsgedächtniskapazität ist bekannt, dass sie als leistungslimitierende Ressource die Performanz bei Indikatoren der fluiden Fähigkeit und der allgemeinen kognitiven Fähigkeit (Colom, Rebollo, Palacios, Juan-Espinosa & Kyllonen, 2004; Engle, Tuholski, Laughlin & Conway, 1999; Kyllonen, 1994; Kyllonen & Christal, 1990; Oberauer, Schulze, Wilhlem & Süß, 2005; Süß, Oberauer, Wittmann, Wilhelm & Schulze, 2002) und auch die Performanz bei mathematischen Fähigkeitsmaßen (Geary & Widaman, 1992; Tirre & Pena, 1993) stark beeinflusst. Daher wird in dieser Arbeit die Position vertreten, dass die Varianz der allgemeinen kognitiven Fähigkeit in Modell 7a in erster Linie interindividuelle Unterschiede in der Arbeitsgedächtniskapazität repräsentiert.

Zusammengenommen können damit die kognitiven Fähigkeiten im Nested-Faktormodell wie folgt interpretiert werden: Die (reliablen Anteile der) stoffgebietspezifischen Residualterme repräsentieren die Verfügbarkeit und Fähigkeit zur Anwendung stoffgebietspezifischen Wissens. Die mathematikspezifische Fähigkeit repräsentiert die stoffgebietsunabhängige Fähigkeit zum mathematischen Problemlösen, und die allgemeine kognitive Fähigkeit repräsentiert die Arbeitsgedächtniskapazität.

Diese Interpretation der generellen mathematischen Fähigkeit und der allgemeinen kognitiven Fähigkeit (in Modell 7a) hilft auch das Ergebnis zu verstehen, dass verbale Fähigkeit keinen Einfluss auf die interindividuellen Unterschiede der stoffgebietspezifischen Fähigkeiten hatte. In der Varianz der mathematikspezifischen Fähigkeit sind die miteinander verwobenen und simultan ablaufenden (siehe hierzu z.B. Hegarty, Mayer & Green, 1992; Hegarty, Mayer & Monk, 1995) Teilprozesse des Textverstehens und des Aufbaus eines mathematischen Problemmodells enthalten. Akzeptiert man diese Interpretation, ist es nicht unmittelbar einleuchtend, weshalb die spezifische verbale Fähigkeit darüber und über die allgemeine kognitive Fähigkeit (als Indikator der Arbeitsgedächtniskapazität) hinaus noch positiven Einfluss auf die Leistung bei Mathematikaufgaben haben sollte.

Zusätzlich zum Verständnis des Zusammenhangs verschiedener psychologischer Konstrukte ist es natürlich mindestens ebenso wichtig, die Institution „Schule“ zu beleuchten, in der Schüler primär ihr mathematisches Wissen und ihre mathematischen Fähigkeiten erwerben (vgl. Geary, 1995; Köller & Baumert, 2002; Köller, Baumert & Schnabel, 2003). Angesichts des zweiten Hauptziels der vorliegenden Arbeit, mathematische Schülerleistung aus der Perspektive eines psychologischen Konstruktes besser zu verstehen, befasst sich das nächste Kapitel mit Schulformunterschieden hinsichtlich der kognitiven Fähigkeiten mathematischer Schülerleistung.

## 8 Studie 2: Schulformunterschiede

Wie unterscheiden sich Schüler, die unterschiedliche Schulformen besuchen, hinsichtlich der kognitiven Fähigkeiten mathematischer Schülerleistung? Diese Frage steht im Zentrum der zweiten Forschungsfrage und wird in zwei Teilfragen untergliedert.

Im Rahmen der ersten Teilfrage werden in Abschnitt 8.1 schulformspezifische Niveauunterschiede der kognitiven Fähigkeiten mathematischer Schülerleistung analysiert. Die zweite Teilfrage geht im Rahmen von Abschnitt 8.2 Fragen der Differenzierung der mathematikspezifischen Fähigkeit nach.

Die Abschnitte zu den beiden Teilfragen sind jeweils wie folgt aufgebaut. Zunächst wird die empirische Befundlage knapp zusammengefasst. Vor dem empirischen Hintergrund und auch mit Blick auf die Überlegungen, die im theoretischen Teil der Arbeit referiert wurden, werden dann die Hypothesen abgeleitet. Anschließend wird die Methodik beschrieben. Daran schließen sich die Darstellung der Ergebnisse sowie eine Diskussion an. Abschließend werden die Befunde der zweiten Forschungsfrage in einen weiteren Kontext eingeordnet und diskutiert.

### 8.1 Niveauunterschiede

#### 8.1.1 Empirische Befundlage und Hypothesen

Die empirische Befundlage zu schulformspezifischen Niveauunterschieden kognitiver Fähigkeiten ist sehr eindeutig: Mit Blick auf die Mittelwerte ergibt sich folgende Rangfolge: Hauptschule < Realschule < Gymnasium. Diese Rangfolge findet man klassenstufenübergreifend sowohl für die fluide Fähigkeit (siehe Tab. 14; Heller & Perleth, 2000) als auch für die generelle mathematische Fähigkeit (siehe Tab. 15; Blum u.a., 2004; Hosenfeld, Köller & Baumert, 1999; Klieme, Neubrand & Lüdtke, 2001; Köller & Baumert, 2002; Kunter, 2005).

Tabelle 14: Schulformspezifische Mittelwertunterschiede der allgemeinen kognitiven Fähigkeit in Effektgrößen  $d$

Klassenstufe	Effektgröße $d$		
	Hauptschule versus Realschule	Hauptschule versus Gymnasium	Realschule versus Gymnasium
5. Klasse	1,23	1,83	0,60
6. Klasse	1,23	1,67	0,44
7. Klasse	1,28	2,05	0,77
8. Klasse	1,30	2,07	0,76
9. Klasse	1,19	2,23	1,04

Die Mittelwertunterschiede basieren auf den Tabellen des KFT-Manuals (Heller & Perleth, 2000, Abschnitt 1.7). Die gepoolte Standardabweichung wurde analog zu Formel 3 (Abschnitt 8.1.3) berechnet. Effektgrößen sind gemittelt über beide Testformen.

Tabelle 15: Schulformspezifische Mittelwertunterschiede der generellen mathematischen Fähigkeit (indiziert durch die globale Leistung in mathematischen Schülerleistungstests) in Effektgrößen  $d$

Klassenstufe/ Altersgruppe	Effektgröße $d$		
	Hauptschule versus Realschule	Hauptschule versus Gymnasium	Realschule versus Gymnasium
7. Klasse <sup>1</sup>	0,30	1,68	1,40
7. Klasse und 8. Klasse <sup>2</sup>	0,54	1,34	0,80
10. Klasse <sup>1</sup>	0,80	2,33	1,54
15-Jährige <sup>3</sup>	0,92	1,75	0,83
15-Jährige <sup>4</sup>	0,89	1,75	0,86

<sup>1</sup> Hosenfeld u.a. (1999, S. 149) auf Grundlage der BIJU-Daten. Es wurde ein mittleres  $d$  berechnet, indem die geschlechterspezifischen Effektgrößen proportional zu den schulformspezifischen Geschlechterverteilungen gewichtet wurden. Die gepoolte Standardabweichung wurde analog zur Formel 3 (Abschnitt 8.1.3) berechnet.

<sup>2</sup> Hosenfeld u.a. (1999, S. 153) auf Grundlage der TIMSS-Daten.  $d$  wurde berechnet indem die schulformspezifischen Mittelwertunterschiede durch die Standardabweichung in der Gesamtstichprobe geteilt wurden.

<sup>3</sup> Klieme, Neubrand u.a. (2001, S. 180) auf Grundlage der PISA-2000-Daten.  $d$  wurde berechnet, indem die schulformspezifischen Mittelwertunterschiede durch die Standardabweichung in der Gesamtstichprobe geteilt wurden.

<sup>4</sup> Blum u.a. (2004, S. 68) auf Grundlage der PISA-2003-Daten. Zunächst wurden getrennt für die Arten des mathematischen Arbeitens die schulformspezifischen Mittelwertunterschiede durch die Standardabweichung in der Gesamtstichprobe geteilt. Anschließend wurde ein mittleres  $d$  über diese separat berechneten Effektgrößen ermittelt. Dieses  $d$  ist in Tabelle 15 eingetragen.

Ingesamt gesehen sind die in Tabelle 14 und 15 eingetragenen Schulformunterschiede – von einigen wenigen Ausnahmen abgesehen – sensu Cohen (1992) als „große Effekte“ (mit  $d$  größer .80) zu beurteilen. Einschränkend ist bei der Beurteilung der berechneten Effektgrößen aber anzumerken, dass diese nur dann vergleichbar sind, wenn sie in gleicher Weise berechnet wurden. Vergleiche sind möglich für alle Effektgrößen innerhalb von Tabelle 14 und für den Vergleich der Effektgrößen von Tabelle 14 mit den mit „1“ gekennzeichneten Studien in Tabelle 15. Weiterhin können in Tabelle 15 die Effektgrößen der mit „2“, „3“ und „4“ gekennzeichneten Studien miteinander verglichen werden. Der Grund für diese Einschränkung ist, dass in den referierten Arbeiten nicht genügend Informationen enthalten waren, um eine gemeinsame Effektstärkenmetrik zu definieren. Dennoch geben die in Tabelle 14 und 15 dokumentierten schulformspezifischen Mittelwertunterschiede einen Eindruck über die tatsächliche Größe der Unterschiede, der deutlich über eine rein qualitative Beschreibung der schulformspezifischen Niveauunterschiede hinausgeht.

Weiterhin ist bei den Ergebnissen von BIJU, TIMSS und PISA zu beachten, dass diese Schulformunterschiede auf einem Modell mathematischer Schülerleistung basieren, das analog zum Standardmodell spezifiziert wurde. Folgt man den Annahmen des Nested-Faktormodells, spiegeln die dokumentierten Schulformunterschiede Unterschiede der allgemeinen kognitiven Fähigkeit *und* der mathematikspezifischen Fähigkeit zwischen den Schulformen wider.

Dies hat direkte Implikationen für die zu erwartenden schulformspezifischen Niveauunterschiede der kognitiven Fähigkeiten mathematischer Schülerleistung. Um die Befunde früherer Studien zu replizieren, wird hier einerseits die Konzeptualisierung der kognitiven Fähigkeiten mathematischer Schülerleistung in Form des Standardmodells (siehe Abb. 16, Modell 5a) gewählt. Anderer-



seits wurde auch die Konzeptualisierung in Form des Nested-Faktormodells (siehe Abb. 16, Modell 7a) untersucht, da damit die mathematikspezifische Fähigkeit getrennt von der allgemeinen kognitiven Fähigkeit analysiert werden kann.

Einschränkend ist vorwegzunehmen, dass stoffgebietspezifische Fähigkeiten wie auch die Konzeptualisierung in Form des *Higher-Order*-Modells (siehe Abb. 16, Modell 6a) in den Analysen in dieser Arbeit nicht berücksichtigt wurden. Denn als Analyseverfahren für die zweite Forschungsfrage empfiehlt sich die konfirmatorische Mehrgruppen-Faktorenanalyse (Carlstedt, 2001; Little, 1997; Lubke, Dolan, Kelderman & Mellenbergh, 2003). Bei Verwendung dieser Methode müssen die Mittelwerte der stoffgebietspezifischen Residualterme auf Null fixiert werden, um das Modell zu identifizieren. Damit ist es ausgeschlossen, Mittelwertunterschiede der stoffgebietspezifischen Fähigkeiten zu untersuchen. Darüber hinaus war aufgrund von Identifikationsproblemen das *Higher-Order*-Faktorenmodell in dieser Arbeit weniger geeignet als das Standardmodell oder das Nested-Faktormodell, da für die residualisierten Faktoren erster Ordnung (z.B.  $M'_{HO}$ ) keine schulformspezifischen Parameter geschätzt werden konnten.

Trotz dieser Einschränkungen werden die Ergebnisse zu den schulformspezifischen Mittelwertunterschieden helfen, ein detailliertes Bild von Schulformunterschieden zu zeichnen. Zunächst werden hierzu die Hypothesen hinsichtlich der mathematischen Fähigkeiten formuliert:

*Hypothese 1a:* Hinsichtlich der mittleren generellen mathematischen Fähigkeit ist folgende Rangfolge der Schulformen zu erwarten: Hauptschule < Realschule < Gymnasium ( $\mu_{M, HS} < \mu_{M, RS} < \mu_{M, HS}$ ). Für diese Hypothese sprechen zwei Gründe:

- Erstens wurde diese Rangfolge in mehreren früheren Studien gefunden (siehe oben).
- Zweitens nehmen leistungsförderliche Aspekte des Mathematikunterrichts, die in erster Linie die mathematischen Fähigkeiten beeinflussen sollten, ebenfalls in dieser Rangfolge zu (vgl. Abschnitt 5.3.2): Pointiert formuliert wird an Gymnasien kognitiv herausfordernd und störungspräventiv unterrichtet. An Hauptschulen hingegen zeichnet sich der Mathematikunterricht durch geringere kognitive Herausforderung und eine erhöhte Anfälligkeit gegenüber Störungen aus. Der Mathematikunterricht an Realschulen liegt zwischen diesen beiden Polen.<sup>25</sup>

*Hypothese 1b:* Aufgrund der beiden oben genannten Gründe wird auch angenommen, dass sich die Schulformen hinsichtlich ihrer mittleren mathematikspezifischen Fähigkeit unterscheiden. Folgende Rangfolge der Schulformen ist zu erwarten: Hauptschule < Realschule < Gymnasium ( $\mu_{M', HS} < \mu_{M', RS} < \mu_{M', HS}$ ).

*Hypothese 1c:* Zusätzlich zur mathematikspezifischen Fähigkeit erklärt die allgemeine kognitive Fähigkeit zu großen Teilen interindividuelle Unterschiede bei Maßen mathematischer Schülerleistung. Zwei Aspekte sind hierbei relevant:

- Erstens ist es ein empirisch oftmals replizierter Befund, dass Schüler höherer Schulformen ein höheres Niveau allgemeiner kognitiver Fähigkeiten aufweisen (z.B. Heller & Perleth, 2000).

---

<sup>25</sup> Individualisierung und Schülerorientierung, wie man sie verstärkt an Hauptschulen findet, haben in der Regel eine positive Wirkung auf motivationale Zielkriterien des Unterrichts und tendenziell keine oder sogar eine negative Wirkung auf kognitive Zielkriterien (Clausen, 2002; Klieme, Schümer & Knoll, 2001; Kunter, 2005).

- Zweitens ist das Niveau allgemeiner kognitiver Fähigkeit ein entscheidender Prädiktor für den Übergang an höhere Schulformen (Baumert, Trautwein & Artelt, 2003).

Daher ist folgende Rangfolge der mittleren allgemeinen kognitiven Fähigkeit zu erwarten: Hauptschule < Realschule < Gymnasium ( $\mu_{g, HS} < \mu_{g, RS} < \mu_{g, HS}$ ).

*Hypothese 1d:* Beim Vergleich der schulformspezifischen Mittelwertunterschiede (gemessen durch Effektgröße  $d$ ) sind größere Unterschiede bei der generellen mathematischen Fähigkeit als bei der mathematikspezifischen Fähigkeit zu erwarten ( $d_{M(HS\ vs.\ RS)} > d_{M'(HS\ vs.\ RS)}$ ;  $d_{M(HS\ vs.\ GY)} > d_{M'(HS\ vs.\ GY)}$ ;  $d_{M(RS\ vs.\ GY)} > d_{M'(RS\ vs.\ GY)}$ ). Der Grund hierfür ist, dass Unterschiede in der generellen mathematischen Fähigkeit schulformspezifische Unterschiede der allgemeinen kognitiven Fähigkeit *und* der mathematikspezifischen Fähigkeit widerspiegeln. Schüler höherer Schulformen sollten sowohl höhere Werte bei der allgemeinen kognitiven Fähigkeit als auch bei der mathematikspezifischen kognitiven Fähigkeit erzielen. Folgt man den Annahmen des Nested-Faktormodells, so ist offensichtlich, dass sich schulformspezifische Mittelwertunterschiede in beiden kognitiven Fähigkeiten addieren. Dies ist bei der mathematikspezifischen Fähigkeit nicht der Fall, da sie unabhängig von der allgemeinen kognitiven Fähigkeit ist (beide Faktoren stehen orthogonal zueinander). Daher sind für die mathematikspezifische Fähigkeit auch geringere Schulformunterschiede zu erwarten.

*Hypothese 1e:* Aufgrund der postulierten Äquivalenz von fluider Fähigkeit und allgemeiner kognitiver Fähigkeit (Gustafsson, 1984) sollten Schulformunterschiede der beiden kognitiven Fähigkeiten gleich stark ausgeprägt sein ( $d_{Gf(HS\ vs.\ RS)} = d_g(HS\ vs.\ RS)$ ;  $d_{Gf(HS\ vs.\ GY)} = d_g(HS\ vs.\ GY)$ ;  $d_{Gf(RS\ vs.\ GY)} = d_g(RS\ vs.\ GY)$ ).

### 8.1.2 Methode

*Stichprobe.* Zur Prüfung der Hypothesen zu schulformspezifischen Niveauunterschieden wurden Hauptschulen, Realschulen und Gymnasien als Schulformen ausgewählt, da bei diesen Schulformen die institutionelle Trennung mit der Leistungsgruppierung eindeutig einherging. Ausgehend von der Analysestichprobe, die auch der ersten Forschungsfrage zu Grunde lag (siehe Abschnitt 7.1.2 und Tab. 3), wurde eine Substichprobe von Neuntklässlern dieser drei Schulformen gebildet. Insgesamt bestand die verwendete Stichprobe aus 4.491 Hauptschülern, 7.077 Realschülern sowie 8.908 Gymnasiasten.

*Manifeste Variablen.* Manifeste Variablen waren die WLE-Scores, die bei der ersten Forschungsfrage verwendet wurden: Es standen die vier stoffgebietspezifischen WLE-Scores als Maße der mathematischen Schülerleistung, die beiden WLE-Scores zur Messung fluider Fähigkeit sowie die drei WLE-Scores zur Messung der verbalen Fähigkeit (indiziert durch die PISA-Lesefähigkeit) zur Verfügung. Die schulformspezifischen deskriptiven Statistiken der manifesten Variablen sind im Anhang dokumentiert.

*Invarianz bei Mehrgruppen-Faktorenanalysen.* Wie können die Hypothesen zu Mittelwertunterschieden kognitiver Fähigkeiten statistisch überprüft werden? Die zentrale Anforderung an das statistische Modell ist, dass „Äpfel mit Äpfel und nicht mit Birnen“ (Horn & McArdle, 1992) verglichen werden. So soll zum Beispiel der Faktor, der die mathematikspezifische Fähigkeit von Hauptschülern repräsentiert, vergleichbar sein mit dem Faktor, der die mathematikspezifische Fähigkeit von Gymnasiasten repräsentiert.

Die Zulässigkeit des Vergleichs kann mithilfe *konfirmatorischer Mehrgruppen-Faktorenanalysen* geprüft werden (Little, 1997; Lubke u.a., 2003). Dabei werden sukzessive Parameterrestriktionen (*constraints*) der (unstandardisierten) Intercepts  $\mu$  und der (unstandardisierten) Faktorladungen  $\lambda^*$  der manifesten Variablen gesetzt, um diese Parameter über die Gruppen hinweg *invariant* zu setzen.<sup>26</sup> Die Zulässigkeit der Vergleiche von Statistiken latenter Variablen ist an bestimmte Parameterrestriktionen geknüpft:

- (a) *Konfigurale Invarianz* (Horn & McArdle, 1992). Bei dieser Form der Invarianz laden über die Gruppen hinweg die manifesten Variablen auf denselben latenten Variablen, und das Muster der Null-Ladungen ist für alle Gruppen gleich. Die Faktorladungen, die Residualvarianzen und die Mittelwerte der manifesten Variablen können (innerhalb und) zwischen den Gruppen variieren, das heißt, diese Modellparameter werden *nicht* invariant über die Gruppen gesetzt.

Wenn die Bedingung konfiguraler Invarianz erfüllt ist, ist jedoch der Vergleich von Faktorvarianzen und -kovarianzen sowie von latenten Mittelwerten *nicht* möglich. Hierfür müssen weitere Bedingungen erfüllt sein.

- (b) *Metrische Invarianz* (Horn & McArdle, 1992). Bei der metrischen Invarianz sind zusätzlich zu den Bedingungen der konfiguralen Invarianz die unstandardisierten Ladungen der manifesten Variablen invariant über die k Gruppen. Es gilt also:

$$\lambda^*_{i1} = \dots = \lambda^*_{ik} \quad (1)$$

Wenn die Bedingung der metrischen Invarianz erfüllt ist, ist die Bedeutung eines Faktors (zur Erklärung interindividueller Unterschiede bei den jeweiligen manifesten Variablen) in allen Gruppen gleich. Somit können Faktorvarianzen und -kovarianzen zwischen den Gruppen verglichen werden.

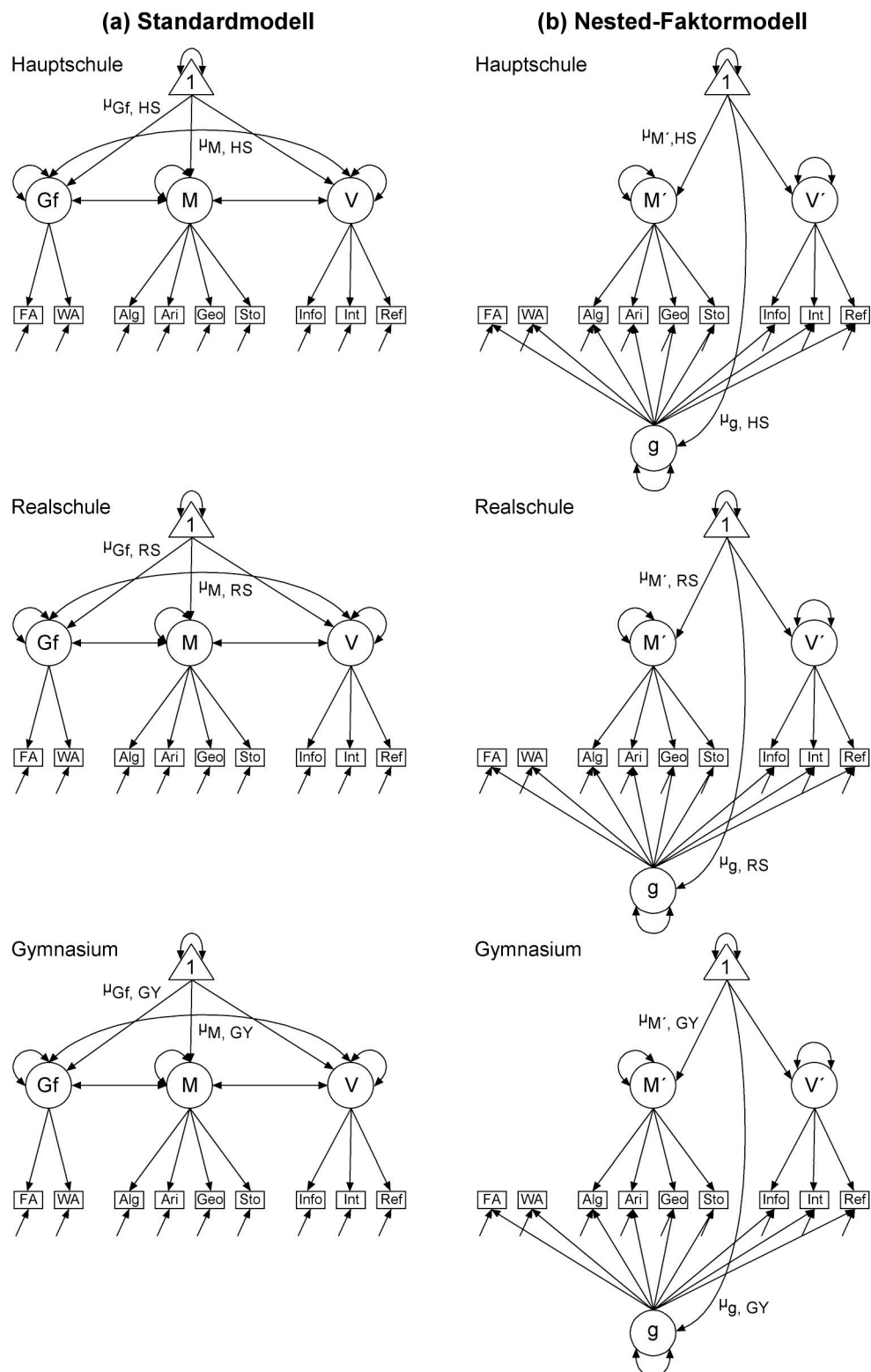
- (c) *Skalare Invarianz* (Millsap & Kwok, 2004). Möchte man darüber hinaus noch die latenten Mittelwerte miteinander vergleichen, muss zum einen die metrische Invarianzbedingung erfüllt sein. Zum anderen müssen zusätzlich die Intercepts der manifesten Variablen über die Gruppen hinweg invariant sein:

$$\mu_i = \dots = \mu_k \quad (2)$$

*Modellspezifikation.* Zur Prüfung der Hypothesen 1a bis 1e wurden das Standardmodell (Abb. 18a) und das Nested-Faktormodell (Abb. 18b) als Mehrgruppen-Faktorenmodell mit den Schulformen Hauptschule, Realschule und Gymnasium als Gruppierungsvariable spezifiziert. Dabei wurden sukzessive die Parameter*constraints* zur Prüfung konfiguraler Invarianz, metrischer Invarianz und der skalaren Invarianz gesetzt. Zur Identifikation der Faktoren wurde die unstandardisierte Faktorladung von jeweils einer manifesten Variablen auf 1,0 fixiert (siehe auch Tab. 14). Im Standardmodell wurden zur Identifikation des Faktors Gf darüber hinaus die unstandardisierten Faktorladungen von FA und WA auf 1,0 fixiert. Zur Identifikation der latenten Mittelwerte wurde der Mittelwert der latenten Variablen in der Hauptschule auf Null fixiert.

---

<sup>26</sup> Auf die Restriktion gruppenübergreifender gleicher unstandardisierter Residualvarianzen (siehe zu den damit verbundenen Interpretationsmöglichkeiten z.B. Lubke u.a., 2003) wird an dieser Stelle nicht weiter eingegangen, da diese zur Prüfung der Hypothesen in der vorliegenden Arbeit nicht notwendig war.



Gf = fluide Fähigkeit, M = generelle mathematische Fähigkeit, V = verbale Fähigkeit, M' = mathematik-spezifische Fähigkeit, V' = spezifische verbale Fähigkeit, g = allgemeine kognitive Fähigkeit, FA = Figurenanalogien, WA = Wortanalogien, Alg = Algebra, Ari = Arithmetik, Geo = Geometrie, Sto = Stochastik, Info = Informationen ermitteln, Int = textbezogenes Interpretieren, Ref = Reflektieren und Bewerten. Intercepts (dies entspricht einem Pfeil von der Dreieck-Mittelwertvariable auf die manifesten Variablen) sind aus Gründen der Übersichtlichkeit nicht eingezeichnet.

Abbildung 18: Modelle zur Analyse von Schulformunterschieden

*Schätzen der Modellparameter.* Zur Schätzung der Modellparameter wurde der Schätzalgorithmus MLR und das Modul „complex“ verwendet, die in Mplus 3.01 (Muthén & Muthén, 1998–2004b) implementiert sind und simultan das Problem fehlender Werte und einer genesteten Datenstruktur berücksichtigen (vgl. Abschnitt 7.2.2).

*Evaluierung des Modell-Fits.* Wie auch bei der herkömmlichen konfirmatorischen Faktorenanalyse konnte der Modell-Fit anhand des inferenzstatistischen  $\chi^2$ -*Goodness-of-Fit*-Tests und deskriptiver Fit-Indizes (z.B. CFI, RMSEA, SRMR) bewertet werden.

Zur Prüfung, ob die verschiedenen Formen der Invarianzannahmen erfüllt sind, wurden Varianten des Standardmodells und des Nested-Faktormodells miteinander verglichen, in denen sukzessive die jeweiligen Parameter*constraints* eingeführt wurden. Hierfür bot sich folgende Reihenfolge an: (1) konfigurale Invarianz, (2) metrische Invarianz und (3) skalare Invarianz (vgl. Little, 1997).

Ob die zusätzlichen Parameter*constraints* zu einer signifikanten Verschlechterung des Modell-Fits führten, konnte anhand der Differenzen der  $\chi^2$ -*Goodness-of-Fit*-Tests ermittelt werden. Allerdings ist hierbei zu bedenken, dass dieser Differenzentest sehr sensitiv gegenüber großen Stichproben und der Anzahl restringierter Parameter ist (Little, 1997). Daher schlägt Little (1997) mehrere zusätzliche Kriterien zur Bewertung der Invarianzannahmen vor. Dazu gehören:

- Der globale Modell-Fit gemessen an den deskriptiven Fit-Statistiken ist akzeptabel.
- Nach Little ist eine marginale Verschlechterung deskriptiver Fit-Indizes bei Hinzufügen von Parameter*constraints* zur Prüfung einer Invarianzannahme zu tolerieren. So sind für ihn beispielsweise Verschlechterungen des *Non Normed Fit Index* (NNFI) bis zu .05 noch akzeptabel.
- Das letztlich favorisierte Modell ist inhaltlich bedeutungsvoller und sparsamer (gemessen an der Anzahl frei geschätzter Parameter) als die alternativen Modelle.

*Inferenzstatistische Prüfung von Hypothesen.* Im Gegensatz zur Prüfung der Invarianzannahmen schlägt Little (1997) vor, die Hypothesen zu latenten Mittelwert-, Varianz- und Kovarianzunterschieden zwischen den Gruppen inferenzstatistisch mit  $\chi^2$ -Differenzentests abzusichern. Die Rationale hierfür ist nach Little, dass die Hypothesen zu Gruppenunterschieden theoretisch abgeleitet werden und daher auch inferenzstatistisch geprüft werden sollten.

In der vorliegenden Arbeit wird der Rationale von Little zur Bewertung der Invarianzannahmen sowie zum Prüfen der Hypothesen gefolgt. Bei inferenzstatistischen Hypothesentests wurde ein Signifikanzniveau von  $p < .05$  gesetzt.

### 8.1.3 Ergebnisse

*Evaluierung des Modell-Fits.* Unter allen drei Invarianzbedingungen wiesen – gemessen an den *Cutoff*-Werten zur Bewertung deskriptiver Fit-Indizes (vgl. Abschnitt 7.2.2) – sowohl das Standardmodell wie auch das Nested-Faktormodell guten Modell-Fit auf (Tab. 16).

Tabelle 16: Globale Modell-Fit-Indizes unter verschiedenen Invarianzbedingungen

Modell	$\chi^2$	df	SF	CFI	RMSEA	SRMR
<i>Standardmodell</i>						
Konfigurale Invarianz	862	77	1.028	0,99	0,04	0,03
Metrische Invarianz	1.411	85	1.003	0,98	0,05	0,04
Skalare Invarianz	2.715	97	1.037	0,95	0,06	0,06
<i>Nested-Faktormodell</i>						
Konfigurale Invarianz	624	60	0.992	0,99	0,04	0,02
Metrische Invarianz	1.392	86	1.013	0,98	0,05	0,04
Skalare Invarianz	2.230	98	1.016	0,96	0,06	0,05

Alle  $\chi^2$ -*Goodness-of-Fit*-Statistiken waren signifikant ( $p < .05$ ).

Zwar führte die Einführung der Parameter*constraints* zu signifikanten Verschlechterungen des Modell-Fits (Tab. 17), doch diese Verschlechterungen konnten als eher marginal betrachtet werden. Die stärkste „Verschlechterung“ resultierte (gerundet) mit .03 für den CFI im Standardmodell aus dem Setzen der Mittelwertrestriktionen für die skalare Invarianz (siehe Tab. 16 und Tab. 17). Dennoch war der Modell-Fit des skalar invarianten Standardmodells immer noch als akzeptabel anzusehen.

Tabelle 17: Veränderung der globalen Modell-Fit-Indizes bei Einführung der Parameter*constraints* zur Prüfung der Invarianz

Modellvergleich	$\Delta\chi^2$	$\Delta df$	$\Delta CFI$	$\Delta RMSEA$	$\Delta SRMR$
<i>Standardmodell</i>					
Konfigurale Invarianz versus metrische Invarianz	694*	12	0,01	-0,01	-0,01
Metrische Invarianz versus skalare Invarianz	1.096*	12	0,03	-0,02	-0,02
<i>Nested-Faktormodell</i>					
Konfigurale Invarianz versus metrische Invarianz	745*	26	0,02	-0,01	-0,01
Metrische Invarianz versus skalare Invarianz	825*	12	0,01	-0,01	-0,02

\*  $p < .05$ .

Zur Berechnung der  $\chi^2$ -Differenz wurde die Korrekturformel aus dem technischen Anhang von Mplus verwendet (Muthén & Muthén, 1998–2004a, S. 22).

War es zulässig, die latenten Mittelwerte im Rahmen des Standardmodells und des Nested-Faktormodells zu vergleichen? Die Kriterien zur Bewertung des Modell-Fits und die marginale Verschlechterung der deskriptiven Fitindizes (siehe aber auch Cheung & Rensvold, 2002, die aber schon bei einer Differenz des CFI von größer .01 die Invarianzvoraussetzung in Frage stellen) sprachen dafür, die skalarinvarianten „Versionen“ beider Modelle für die weiteren Analysen zu verwenden. Darüber hinaus ermöglicht erst die skalare Invarianz beider Modelle die inhaltlich

bedeutsame Analyse von Schulformunterschieden. Damit waren alle drei von Little (1997) genannten Bedingungen erfüllt.

*Hypothesenprüfung.* Die nachfolgenden Darstellungen beschränken sich nur auf die Modellparameter, die zur Prüfung der Hypothesen relevant waren. Die latenten Mittelwerte der kognitiven Fähigkeiten sind in Tabelle 18 eingetragen. Dabei ist zu beachten, dass bei allen kognitiven Fähigkeiten der Mittelwert der Hauptschüler auf Null fixiert wurde, um die Modelle zu identifizieren. Die latenten Mittelwerte der beiden anderen Schulformen konnten somit als Mittelwertdifferenzen zwischen dem jeweiligen latenten Schulmittelwert und dem latenten Mittelwert der Hauptschüler interpretiert werden. Auf Unterschiede der verbalen Fähigkeiten (V oder V') wird in dieser Arbeit nicht weiter eingegangen, da diese keinen inkrementellen Beitrag zur Erklärung interindividueller Unterschiede mathematischer Schülerleistung leisten konnten (vgl. Abschnitt 7.2, Modell 7b).

Tabelle 18: Schulformspezifische unstandardisierte latente Mittelwerte ( $\mu$ ), korrespondierende 95%-Konfidenzintervalle [95%-KI]

Schulform	$\mu_M/\mu_M'$ [95%-KI]	$\mu_{Gf}/\mu_g$ [95%-KI]	$\mu(V)/\mu(V')$ [95%-KI]
<i>Standardmodell</i>			
Hauptschule	0,00 <sup>a</sup>	0,00 <sup>a</sup>	0,00 <sup>a</sup>
Realschule	0,93 [0,85; 1,00]	0,75 [0,69; 0,80]	0,87 [0,80; 0,93]
Gymnasium	2,21 [2,15; 2,28]	1,57 [1,53; 1,62]	1,69 [1,63; 1,75]
<i>Nested-Faktormodell</i>			
Hauptschule	0,00 <sup>a</sup>	0,00 <sup>a</sup>	0,00 <sup>a</sup>
Realschule	0,26 [0,20; 0,33]	0,67 [0,60; 0,74]	0,33 [0,28; 0,38]
Gymnasium	0,88 [0,79; 0,98]	1,45 [1,37; 1,53]	0,55 [0,49; 0,60]

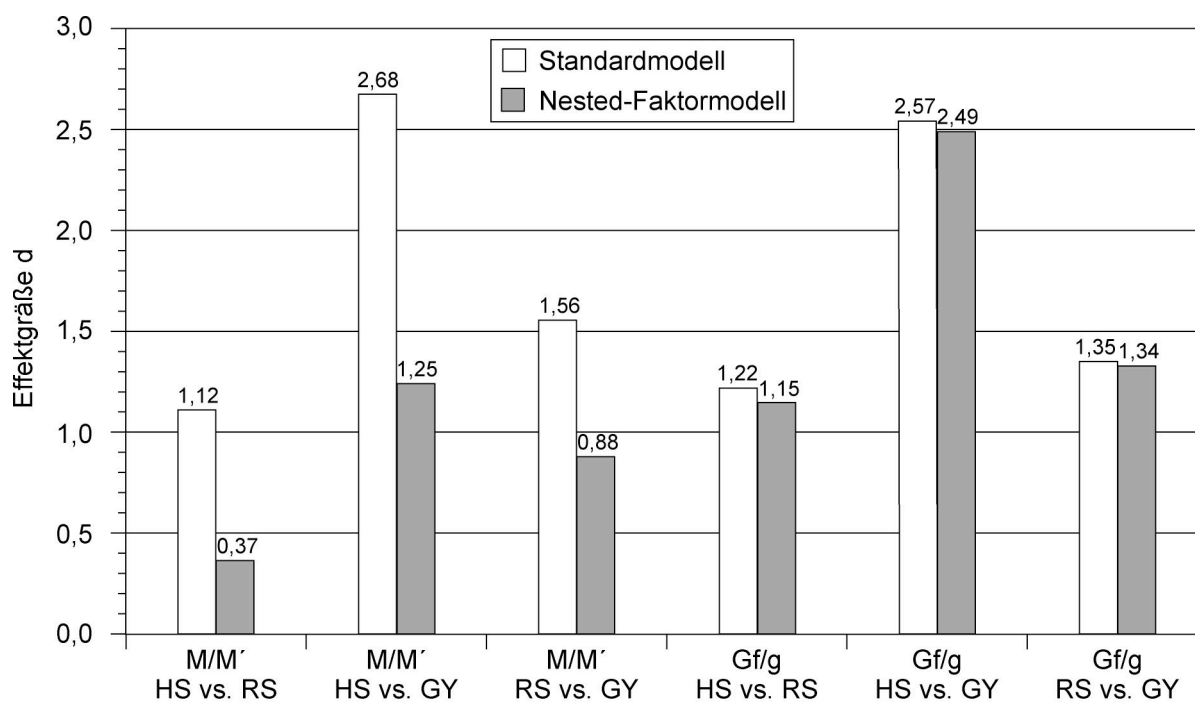
Es ist zu beachten, dass die unstandardisierten Parameter nicht zwischen dem Standardmodell und dem Nested-Faktormodell verglichen werden können.

<sup>a</sup> Diese Parameter wurden auf Null fixiert, um das Modell zu identifizieren. Daher wurde kein Standardfehler geschätzt und somit konnte kein Konfidenzintervall berechnet werden.

Als Effektgröße  $d$  wurde eine Effektgröße analog zu Cohens (1992)  $d$  berechnet. Die zu Grunde liegende gepoolte Standardabweichung  $SD_{pooled}$  wurde folgendermaßen aus den schulformspezifischen Varianzen  $\sigma^2$  berechnet (vgl. Bortz & Döring, 2002, S. 605):

$$SD_{pooled} = \sqrt{\frac{N_{HS} \cdot \sigma_{HS}^2 + N_{RS} \cdot \sigma_{RS}^2 + N_{GY} \cdot \sigma_{GY}^2}{N_{HS} + N_{RS} + N_{GY}}} \quad (3)$$

$N$  stellt dabei die Stichprobengröße der jeweiligen Schulformen Hauptschule (HS), Realschule (RS) und Gymnasium (GY) dar. Die resultierenden Effektgrößen sind in Abbildung 19 eingezeichnet.



M = generelle mathematische Fähigkeit (weißer Balken), M' = mathematikspezifische Fähigkeit (grauer Balken), Gf = fluide Fähigkeit (weißer Balken), g = allgemeine kognitive Fähigkeit (grauer Balken).

Abbildung 19: Effektgrößen der schulformspezifischen Unterschiede in den mathematischen Fähigkeiten

Auf Grundlage der geschätzten latenten Mittelwerte und der berechneten Effektgrößen konnten die postulierten Hypothesen geprüft werden. Zunächst werden die Hypothesen bezüglich der mathematischen Fähigkeiten untersucht.

*Schulformspezifische Unterschiede in den mathematischen Fähigkeiten.* Vergleich man die latenten Mittelwerte der generellen mathematischen Fähigkeit im Standardmodell, so war ersichtlich, dass Hauptschüler im Mittel geringere Leistungen erzielten als Realschüler ( $d = 1,12$ ) oder Gymnasiasten ( $d = 2,68$ ). Auch schnitten Realschüler durchschnittlich schwächer ab als Gymnasiasten ( $d = 1,56$ ).

Zur inferenzstatistischen Überprüfung der Mittelwertunterschiede wurden die Mittelwerte von zwei Schulformen auf den gleichen Wert fixiert. Eine statistisch signifikante Differenz der  $\chi^2$ -Goodness-of-Fit-Werte war indikativ für einen überzufälligen Mittelwertsunterschied. Zur Berechnung der  $\chi^2$ -Differenzen wurde die empfohlene Korrekturformel (Muthén & Muthén, 1998–2004a, S. 22) verwendet. Bei der inferenzstatistischen Überprüfung war die mittlere Leistung von Hauptschülern und Gymnasiasten bei der generellen mathematischen Fähigkeit statistisch signifikant verschieden (Tab. 19). Bei den Vergleichen der beiden anderen Schulformvergleiche resultierten zwei  $\chi^2$ -Differenzen kleiner Null, die nicht interpretiert werden konnten. Nach Satorra und Bentler (2001, S. 511) ist dies entweder indikativ dafür, dass die Stichprobe zu klein ist (was in der vorliegenden Arbeit ausgeschlossen werden kann) oder dass das restringiertere Modell zu stark vom weniger restringierten Modell abweicht. Letzteres schien der Fall zu sein, denn der  $\chi^2$ -Goodness-of-Fit-Wert des restringierten Modells verschlechterte sich sehr deutlich im Vergleich zum unrestringierten Modell. Dies zeigte ein deskriptiver Vergleich der Werte für die skalarinvarianten Modelle aus Tabelle 16 und den restringierten Modellen in Tabelle 19. So verschlechterte



sich beispielsweise der  $\chi^2$ -Goodness-of-Fit-Wert im Standardmodell von  $\chi^2 = 2.715$  auf  $\chi^2 = 5.127$  beim Test für die Gleichheit der Mittelwerte der generellen mathematischen Fähigkeit von Haupt- und Realschülern. Unabhängig von den Problemen mit den  $\chi^2$ -Differenzentests wiesen die Konfidenzintervalle der schulformspezifischen Mittelwerte, die sich nicht überlappten, darauf hin, dass die jeweiligen Mittelwerte erwartungskonform statistisch signifikant voneinander verschieden waren. Insgesamt gesehen wurde Hypothese 1a somit von den empirischen Ergebnissen gestützt.

Tabelle 19: Test auf schulformspezifische Niveauunterschiede der kognitiven Fähigkeiten mathematischer Schülerleistung

Modellvergleich	$\chi^2$	df	SF	CFI	RMSEA	SRMR	$\Delta\chi^2$
<i>Standardmodell</i>							
$\mu_M(\text{HS}) = \mu_M(\text{RS})$	5.127	98	1.021	0,91	0,09	0,11	-4.555
$\mu_M(\text{HS}) = \mu_M(\text{GY})$	10.816	98	1.028	0,80	0,13	0,42	53.568
$\mu_M(\text{RS}) = \mu_M(\text{GY})$	8.468	98	0.934	0,84	0,11	0,15	-562
<i>Nested-Faktormodell</i>							
$\mu_M(\text{HS}) = \mu_M(\text{RS})$	2.340	99	1.016	0,96	0,06	0,06	110
$\mu_M(\text{HS}) = \mu_M(\text{GY})$	2.464	99	1.035	0,96	0,06	0,05	98
$\mu_M(\text{RS}) = \mu_M(\text{GY})$	2.514	99	1.023	0,95	0,06	0,06	179
$\mu_g(\text{HS}) = \mu_g(\text{RS})$	4.260	99	1.011	0,92	0,08	0,09	3.916
$\mu_g(\text{HS}) = \mu_g(\text{GY})$	9.064	99	1.021	0,83	0,12	0,28	4.625
$\mu_g(\text{RS}) = \mu_g(\text{GY})$	5.767	99	1.004	0,89	0,09	0,12	-20.486

Die Modell-Fit-Indizes resultierten aus einem Modell, in dem die jeweiligen Mittelwerte auf den gleichen Wert fixiert wurden. Zur Berechnung der  $\chi^2$ -Differenz wurde die Korrekturformel aus dem technischen Anhang von Mplus verwendet (Muthén & Muthén, 1998–2004a, S. 22). Alle  $\chi^2$ -Differenzen werden mit  $df = 1$  getestet. SF = Skalierungsfaktor.

Auch bei der mathematikspezifischen Fähigkeit im Nested-Faktormodell zeigte sich die erwartete Rangfolge der Schulformen. Die mathematikspezifische Fähigkeit von Hauptschülern lag im Mittel um 0,37 Standardabweichungen unter der durchschnittlichen Leistung von Realschülern und 1,25 Standardabweichungen unter der mittleren Leistung von Gymnasiasten. Im Durchschnitt erzielten Realschüler um 0,88 Standardabweichungen schlechtere Leistungen als Gymnasiasten. Wie zu erwarten, waren alle Mittelwertunterschiede der mathematikspezifischen Fähigkeit statistisch signifikant von Null verschieden (Tab. 19). Die deskriptiven Unterschiede wie auch die inferenzstatistischen Analysen stützten somit Hypothese 1b.

Zur Prüfung der Frage, ob die Schulformunterschiede stärker bei der generellen mathematischen Fähigkeit oder der mathematikspezifischen Fähigkeit ausgeprägt waren (Hypothese 1d), war es notwendig, die schulformspezifischen Mittelwertunterschiede beider Modelle auf derselben Metrik zu vergleichen. Hierzu wurden die Effektgrößen  $d$  verwendet. Beim Vergleich der Effektgrößen war offensichtlich, dass die schulformspezifischen Unterschiede in der generellen mathematischen Fähigkeit deutlich stärker ausgeprägt waren als bei der mathematikspezifischen Fähigkeit. Betrachtete man zum Beispiel den Mittelwertunterschied zwischen der Hauptschule und

dem Gymnasium, so betrug dieser 2,68 Standardabweichungen für die generelle mathematische Fähigkeit und 1,25 Standardabweichungen für die mathematikspezifische Fähigkeit (vgl. Abb. 19). Die Differenz der Effektgrößen zwischen den beiden geschätzten Schulformunterschieden betrug also 1,43 Standardabweichungen (2,68–1,25).

Für den Vergleich von Effektgrößen aus einer Einzelstudie gibt es meines Wissens keinen inferenzstatistischen Test. Hypothese 1d wurde somit von den empirischen Ergebnissen gestützt, konnte aber nicht zufallskritisch abgesichert werden.

Zusammenfassend kann man an dieser Stelle festhalten, dass die postulierten Hypothesen hinsichtlich der mathematischen Fähigkeiten von den empirischen Ergebnissen gestützt wurden. Traf dies auch für die Hypothesen zur allgemeinen kognitiven Fähigkeit und der fluiden Fähigkeit zu?

*Schulformspezifische Unterschiede der allgemeinen kognitiven Fähigkeit und der fluiden Fähigkeit.* Schüler an der Hauptschule erzielten im Durchschnitt niedrigere Leistungen bei der allgemeinen kognitiven Fähigkeit als Realschüler ( $d = 1,15$ ) und schnitten durchschnittlich schwächer ab als Gymnasiasten ( $d = 2,49$ ). Ebenso lag die mittlere Leistung von Realschülern unter der von Gymnasiasten ( $d = 1,34$ ).

Bei der inferenzstatistischen Überprüfung der Mittelwertunterschiede waren die mittleren Leistungen von Hauptschülern und Realschülern sowie von Hauptschülern und Gymnasiasten statistisch signifikant verschieden. Jedoch resultierte für den Schulformvergleich der Mittelwerte von Realschülern und Gymnasiasten eine  $\chi^2$ -Differenz kleiner als Null. Dieser negative Wert konnte nicht interpretiert werden (siehe oben). Jedoch indizierten die deutlichen Verschlechterungen der  $\chi^2$ -Goodness-of-Fit-Werte wie auch die Konfidenzintervalle der schulformspezifischen Mittelwerte, die sich nicht überlappten, dass diese Mittelwertdifferenz statistisch signifikant von Null verschieden war. Hypothese 1c wurde somit empirisch gestützt.

Die Ergebnisse stützten auch die Annahme von Hypothese 1e, dass die Schulformunterschiede hinsichtlich der allgemeinen kognitiven Fähigkeit und der fluiden Fähigkeit gleich stark ausgeprägt waren. Ein Vergleich der Effektgrößen zeigte, dass unabhängig vom spezifizierten Modell nahezu identische Werte resultierten: Der größte Unterschied zwischen den Effektgrößen betrug .08 Standardabweichungen für den Schulformunterschied der mittleren Leistungen von Hauptschülern und Gymnasiasten. Ob die minimalen Unterschiede statistisch signifikant waren, konnte aufgrund fehlender inferenzstatistischer Tests nicht geprüft werden.

#### 8.1.4 Diskussion

Die wichtigsten Ergebnisse der vorangegangenen Analysen lassen sich wie folgt zusammenfassen:

- Die skalarinvarianten Versionen des Standardmodells und des Nested-Faktormodells konnten die empirischen Relationen gut approximieren.
- Bei allen analysierten Fähigkeiten zeigte sich hypothesenkonform die erwartete Rangfolge der schulformspezifischen Mittelwerte: Hauptschule < Realschule < Gymnasium.

- Schulformunterschiede in der fluiden Fähigkeit waren in etwa genauso groß wie Schulformunterschiede in der allgemeinen kognitiven Fähigkeit.
- Schulformunterschiede der generellen mathematischen Fähigkeit waren deutlich stärker ausgeprägt als Schulformunterschiede in der mathematikspezifischen Fähigkeit.

Die Hypothesen zu den Niveauunterschieden wurden einerseits auf Grundlage der empirischen Befunde früherer Forschung und andererseits aus Befunden der Unterrichtsforschung abgeleitet. Wie zu erwarten war, waren die Ergebnisse zu schulformspezifischen Niveauunterschieden kognitiver Fähigkeiten sehr eindeutig sowohl bei den mathematischen Fähigkeiten als auch bei der allgemeinen kognitiven Fähigkeit (bzw. fluiden Fähigkeit). Ob diese Unterschiede nun bereits vor dem Übertritt an höhere Schulformen bestanden haben oder ob sie erst durch die unterschiedlichen Unterrichtsmilieus entstanden sind, kann anhand der vorliegenden Daten nicht beantwortet werden. Erst ein längsschnittliches Feldexperiment, in dem eine repräsentative Stichprobe von Schülern zufällig verschiedenen Schulformen zugewiesen wird, könnte hier Klarheit schaffen. Jedoch ist klar, dass dieses Experiment aus ethischen Gründen niemals durchgeführt werden sollte. Wie man dennoch dieses Experiment approximieren kann, wird im Rahmen der Diskussion in Abschnitt 8.3 weiter ausgeführt, wenn methodische Überlegungen zur Analyse von Unterrichtseffekten diskutiert werden.

Ein davon unabhängiger zentraler Befund der vorliegenden Arbeit zeigt, dass die Schulformunterschiede in der generellen mathematischen Fähigkeit wesentlich deutlicher ausgeprägt waren als in der mathematikspezifischen Fähigkeit. Erklärt wurde dies dadurch, dass in die generelle mathematische Fähigkeit Unterschiede der allgemeinen kognitiven Fähigkeit und der mathematikspezifischen Fähigkeit gemischt eingingen. Vor dem Hintergrund, dass in bisherigen Studien (Blum u.a., 2004; Klieme, Neubrand u.a., 2001; Köller & Baumert, 2002; Kunter, 2005) in der Regel die generelle mathematische Fähigkeit betrachtet wurde, wurden die Unterschiede zwischen den Schulformen mit Blick auf die mathematikspezifische Fähigkeit somit tendenziell überschätzt, bzw. die berichteten Unterschiede bezogen sich auf ein unspezifiziertes „Varianzamalgame“ der allgemeinen kognitiven Fähigkeit und der mathematikspezifischen Fähigkeit. In der vorliegenden Arbeit ist es gelungen, dieses Amalgam zu trennen, um so ein differenzierteres Bild der schulformspezifischen Niveauunterschiede zu zeichnen.

Geht man davon aus, dass die allgemeine kognitive Fähigkeit und die mathematikspezifische Fähigkeit bedeutsame Prädiktoren des Lernerfolgs (vgl. Abschnitt 5.1), aber auch Resultat von Lernprozessen sind (Ceci, 1991; Helmke & Weinert, 1997; Köller & Baumert, 2002), haben diese Befunde auch für die These von Schulformen als differenzielle Entwicklungsmilieus (Baumert, Köller & Schnabel, 2000) Bedeutung. Insbesondere die Interpretation der mathematikspezifischen Fähigkeit ist hierbei interessant: Wenn zwei Schüler die gleiche Leistung bei der allgemeinen kognitiven Fähigkeit erzielten, der eine aber die Realschule und der andere die Hauptschule besuchte, hatte der Realschüler eine um 0,37 Standardabweichungen bessere Leistung als der Hauptschüler in der mathematikspezifischen Fähigkeit. Für den Vergleich eines Hauptschülers mit einem Gymnasiasten (gleicher allgemeiner kognitiver Fähigkeit) betrug der Leistungsvorsprung bei der mathematikspezifischen Fähigkeit sogar 1,25 Standardabweichungen. Unter der Annahme, dass der Lernzugewinn in Mathematik in einem Schuljahr bei etwa 0,3 Standardabweichungen liegt (Baumert & Artelt, 2002; Köller, Baumert & Schnabel, 2003), würden die Leistungsunterschiede zwischen Hauptschülern und Realschülern (Gymnasiasten) gleicher allgemeiner kognitiver Fähigkeit einem Unterschied von etwa einem bzw. vier Schuljahren entsprechen.

Folgt man der Annahme, dass Unterschiede in den Eingangsvoraussetzungen unterschiedliche Entwicklungstrajektorien in den mathematischen Fähigkeiten bedingen, dann sollte die Leistungsschere zwischen den Schulformen mit höheren Klassenstufen zunehmen. Dies impliziert eine Zunahme der Heterogenität der Schülerschaft in der Gesamtpopulation der Schüler wie auch innerhalb der Schulformen (siehe hierzu z.B. Köller & Baumert, 2001; siehe aber auch Schneider & Stefanek, 2004). Erste Hinweise auf dieses Phänomen zeigt ein Vergleich der schulformspezifischen Varianzen der mathematikspezifischen Fähigkeit (Tab. 20): Die Varianz nahm von der Hauptschule über die Realschule hin zum Gymnasium zu.

Im nächsten Abschnitt wird diese Zunahme der Heterogenität der Schülerschaft zwischen und innerhalb von Schulformen im Rahmen der Analysen zur Differenzierung der mathematikspezifischen Fähigkeit detailliert betrachtet.

Tabelle 20: Schulformspezifische Varianz ( $\sigma^2$ ) der mathematikspezifischen Fähigkeit (M') und korrespondierendes 95%-Konfidenzintervall [95%-KI]

Schulform	$\sigma^2(M')$	[95%-KI]
Hauptschule	0,05	[-0,04; 0,13]
Realschule	0,46	[0,31; 0,60]
Gymnasium	0,75	[0,50; 1,01]

## 8.2 Differenzierung der mathematikspezifischen Fähigkeit

### 8.2.1 Empirische Befundlage und Hypothesen

Bei den vorangegangenen Analysen wurden schulformspezifische Unterschiede des Niveaus der kognitiven Fähigkeiten mathematischer Schülerleistung analysiert. Im Unterschied hierzu werden im Rahmen der nächsten beiden Hypothesen einerseits die Heterogenität (in Form der Varianz) und andererseits nur die mathematikspezifische Fähigkeit für Schüler unterschiedlicher Fähigkeitsniveaus sowie unterschiedlicher Schulform untersucht. Diese Analysen fokussieren auf die Differenzierung der mathematikspezifischen Fähigkeit.

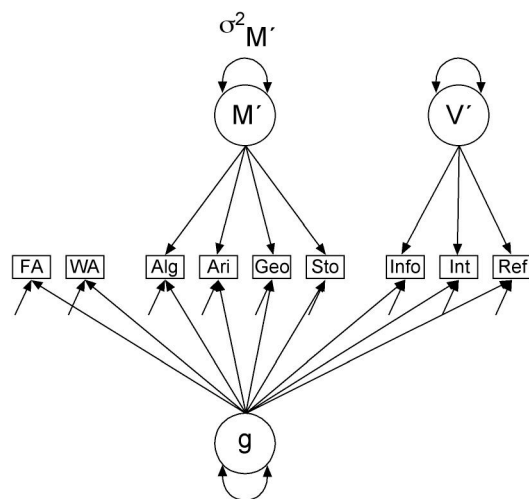
*Generelle Differenzierungshypothese* (Hypothese 2a). Einige Studien weisen darauf hin, dass Personen mit höherem Niveau allgemeiner kognitiver Fähigkeit eine differenziertere Faktorstruktur aufweisen als Personen mit niedrigerem Fähigkeitsniveau (vgl. Abschnitt 4.3.2; Abad, Colom, Juan-Espinosa & Garcia, 2003; Carlstedt, 2001; Deary u.a., 1996; Detterman & Daniel, 1989; Jensen, 2003). In den Studien, die mathematische Fähigkeiten untersuchten (dies sind alle hier zitierten mit Ausnahme der Studie von Carlstedt), nahm die Varianz der mathematikspezifischen Fähigkeitsfaktoren mit zunehmender allgemeiner kognitiver Fähigkeit zu.

Wie in Abschnitt 4.3.2 ausgeführt wurde, kann dieser empirische Befund der Differenzierung mit Jensen (1998) folgendermaßen erklärt werden: Im Zuge der kognitiven Entwicklung wird die allgemeine kognitive Fähigkeit zunehmend in spezialisiertere Fähigkeiten investiert. Mit zunehmender Übung werden diese spezialisierteren Fähigkeiten mehr und mehr automatisiert. Die automatisierbaren Aspekte dieser Fähigkeiten sind dann von der allgemeinen kognitiven Fähig-

keit unabhängig. Folgt man Jensen, dann haben Schüler mit höherer allgemeiner kognitiver Fähigkeit tendenziell die kognitiven Teilprozesse des mathematischen Problemlösens stärker automatisiert. Dies führt (bei einer gleichzeitigen differenziellen Investition in unterschiedliche spezialisierte Fähigkeiten z.B. aufgrund unterschiedlicher fachspezifischer Lernmotivation, siehe hierzu auch Abschnitt 9.6) dazu, dass die Varianz der mathematikspezifischen Fähigkeit in der Gesamtstichprobe der Schüler zunimmt.

Unabhängig von der besuchten Schulform sollte also eine Differenzierung der mathematikspezifischen Fähigkeit zu beobachten sein (vgl. Abb. 20b, Gesamtstichprobe): Je höher das Niveau allgemeiner kognitiver Fähigkeit  $g$  ist, desto größer ist die Varianz der mathematikspezifischen Fähigkeit  $\sigma^2_{M'}$ . Dies bedeutet, dass die Heterogenität der Schülerschaft in der Gesamtstichprobe hinsichtlich der mathematikspezifischen Fähigkeit mit zunehmendem Fähigkeitsniveau zunimmt.

(a) Faktormodell



(b) Schematische Darstellung

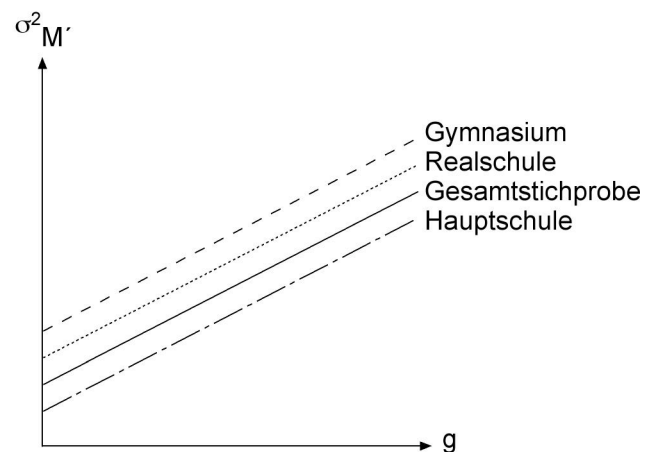


Abbildung 20: Faktormodell und schematische Darstellung der Differenzierung der mathematikspezifischen Fähigkeit ( $M'$ )

*Schulformspezifische Differenzierungshypothese* (Hypothese 2b). Eine weitere Erklärung der Differenzierung stammt von Abad und Kollegen (2003). Abad u.a. erklären die Differenzierung spezifischer Fähigkeiten dadurch, dass Personen mit höherem Fähigkeitsniveau Zugang zu besseren Bildungsmöglichkeiten (z.B. höheren Schulformen) haben. Unter der Annahme, dass an höheren Schulformen leistungsförderliche Aspekte des Mathematikunterrichts stärker ausgeprägt sind (vgl. Abschnitt 5.4; siehe Baumert u.a., 2004; Gruehn, 2000; Klieme, Schümer u.a., 2001; Kunter, 2005), finden Schüler an höheren Schulformen bessere Investitionsmöglichkeiten für ihre allgemeine kognitive Fähigkeit vor (vgl. Abschnitt 4.3.2). Daher sollten Schüler gleichen Fähigkeitsniveaus, die aber unterschiedliche Schulformen besuchen, ein unterschiedliches Ausmaß der Differenzierung aufweisen. Bei gleichem Niveau allgemeiner kognitiver Fähigkeit sollte die Differenzierung der mathematikspezifischen Fähigkeit von Schülern der höheren Schulform stärker ausgeprägt sein (dies entspricht in Abb. 20b den schulformspezifischen Funktionsgeraden zum Zusammenhang von  $g$  und  $\sigma^2_{M'}$ ). Bei gleichem Fähigkeitsniveau sollte also die Varianz der mathematikspezifischen Fähigkeit von Schülern am Gymnasium größer sein als die Varianz von Schülern an der Realschule. Ebenso sollte (bei gleichem Fähigkeitsniveau) die Heterogenität in

Bezug auf die mathematikspezifische Fähigkeit von Realschülern größer sein als die Heterogenität von Hauptschülern. Diese Hypothese ist meines Wissens bislang noch nicht geprüft worden.

Erste Evidenz für diese Hypothese zeigten die schulformspezifischen Unterschiede in der Varianz der mathematikspezifischen Fähigkeit, die in Tabelle 20 dokumentiert sind. Jedoch ist aus Unterschieden der Heterogenität in Tabelle 20 nicht ersichtlich, ob diese Varianzunterschiede sich gleichmäßig über das gesamte Spektrum der allgemeinen kognitiven Fähigkeit verteilen oder nur für bestimmte Fähigkeitsniveaus zu finden sind. Diese Frage wird in den Analysen zur schulformspezifischen Differenzierungshypothese detailliert untersucht.

### 8.2.2 Methode

Bei den Differenzierungshypothesen werden Unterschiede in der Heterogenität der mathematikspezifischen Fähigkeit in Abhängigkeit vom Fähigkeitsniveau (Hypothese 2a) sowie in Abhängigkeit vom Fähigkeitsniveau *und* der Schulform (Hypothese 2b) untersucht. Für die Analysen zur Differenzierungshypothese existiert keine Standardprozedur. Carlstedt (2001) hat eine elegante vierschrittige Methode vorgeschlagen, die in dieser Arbeit verwendet wurde (für eine alternative Vorgehensweise, die allerdings das Problem der Invarianz der spezifischen Fähigkeit über die Gruppen nicht adressiert, siehe Deary u.a., 1996).

Nachfolgend werden die vier Schritte erklärt. Dabei werden die zu Grunde liegenden Stichproben bei der Darstellung des zweiten Schritts, die verwendeten manifesten Variablen, die spezifizierten Modelle und die inferenzstatistischen Tests bei der Darstellung des dritten Schritts beschrieben.

(1) *Schätzung von Faktorscores allgemeiner kognitiver Fähigkeit.* Zunächst wurden für das Nested-Faktormodell (Modell 7a) mit Lisrel 8.51 (Jöreskog, Sörbom, Du Toit & Du Toit, 1999) auf Basis der gesamten Analysetichprobe von 29.386 Schülern Faktorscores für die allgemeine kognitive Fähigkeit geschätzt. Diese Faktorscores hatten gegenüber regressionsanalytisch abgeleiteten Faktorscores den Vorteil, dass die korrelative Struktur (hier die Orthogonalität) der latenten Variablen auch in den Faktorscores erhalten blieb (siehe hierzu auch Grice, 2001).

Einige wenige Schüler (155 von 29.386, dies entspricht 0,5%, siehe auch Anhang) hatten fehlende Werte bei einer oder mehreren der neun manifesten Variablen zur Messung kognitiver Fähigkeiten. Damit die Analysetichprobe in dieser Arbeit über die Forschungsfragen hinweg konstant blieb, wurden für diese Schüler zunächst mit dem Programm NORM (Schafer, 2000) EM-imputierte Werte als beste Punktschätzer (Graham, Cumsille & Elek-Fisk, 2003) der manifesten WLE-Scores bestimmt. Die Faktorscores dieser Schüler und ihre Zuordnung zu den Fähigkeitsgruppen basierten auf den EM-imputierten Werten. Die Modellparameter (siehe unten) basieren jedoch auf *Full Information Maximum Likelihood* Schätzungen auf Grundlage der Daten, in denen die Werte *nicht* imputiert wurden, sondern als fehlend codiert waren.

Diesem Vorgehen wurde gegenüber alternativen Vorgehensweisen zur Behandlung fehlender Werte der Vorzug gegeben:

- Gegenüber listenweisem Fallausschluss lag der Vorteil darin, dass die Stichprobe vergleichbar zur Stichprobe der ersten und dritten Forschungsfrage blieb und durch die Verwendung von *Full Information Maximum Likelihood* unverzerrte Schätzungen von Modellparametern und inferenzstatistischen Prüfgrößen vorlagen.

- Beim *Multiple-Imputation*-Ansatz (z.B. Graham u.a., 2003; Schafer & Graham, 2002) werden mehrere Datensätze generiert, in denen die fehlenden Daten auf Grundlage geschätzter Populationsstatistiken (Mittelwerte, Varianzen und Kovarianzen) der manifesten Variablen ersetzt werden. Demgegenüber hat das in der vorliegenden Arbeit gewählte Verfahren den Vorteil, dass für *multiple imputation* noch keine Methode entwickelt wurde, wie  $\chi^2$ -Differenzen über mehrere imputierte Datensätze hinweg zu integrieren sind. Die  $\chi^2$ -Differenzen waren aber hilfreich, um die Hypothesen 2a und 2b zu prüfen.

(2) *Zuordnung der Schüler zu Fähigkeitsgruppen.* Auf Grundlage der Daten der gesamten Analysestichprobe (N = 29.386) wurden die Faktorscores allgemeiner kognitiver Fähigkeit dazu verwendet, die Schüler in 8, 16 und 32 gleich große *Fähigkeitsgruppen* einzuteilen (für die Wahl dieser Gruppenanzahlen siehe Carlstedt, 2001). Anhand der Fähigkeitsgruppen wurde auf Grundlage der Analysestichprobe die generelle Differenzierungshypothese (Hypothese 2a) geprüft.

Für die Analysen zur schulformspezifischen Differenzierungshypothese wurden die Schüler innerhalb einer Fähigkeitsgruppe nochmals entsprechend ihrer Schulformzugehörigkeit eingeteilt. Damit lagen für jede Fähigkeitsgruppe fünf schulformspezifische Untergruppen vor.

Zur Prüfung der schulformspezifischen Differenzierung wurden nur Schüler der Schulformen Hauptschule, Realschule und Gymnasium ausgewählt, weil an diesen Schulformen die institutionelle Trennung eindeutig mit der Leistungsgruppierung einherging. Weiterhin wurden nur schulformspezifische Fähigkeitsgruppen in die Analysen eingeschlossen, die jeweils von mindestens 100 Schülern besetzt waren (vgl. Abb. 21). So wurde sichergestellt, dass die geschätzte schulformspezifische Varianz der mathematikspezifischen Fähigkeiten auf einer ausreichend großen Fallzahl beruhte.

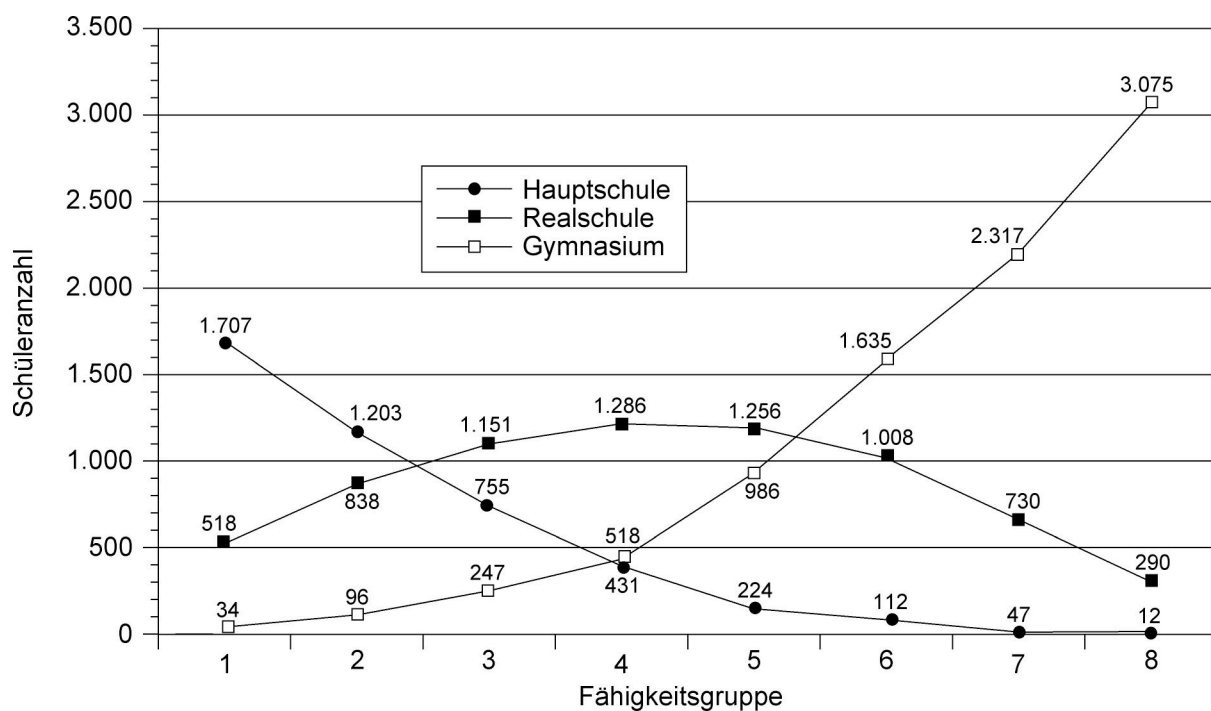


Abbildung 21: Verteilung der Schüler über die Fähigkeitsgruppen in Abhängigkeit von der Schulform

Insgesamt basierten die Analysen zur schulformspezifischen Differenzierungshypothese auf einer Stichprobe von 4.432 Hauptschülern, 7.077 Realschülern sowie 8.778 Gymnasiasten. Daten von insgesamt 191 Schülern (dies waren Hauptschüler aus den Fähigkeitsgruppen 7 und 8 der Hauptschule und Gymnasiasten der Fähigkeitsgruppen 1 und 2) wurden nicht berücksichtigt. Innerhalb der schulformspezifischen Fähigkeitsgruppen war mit Ausnahme der Extremgruppen das schulformspezifische Niveau der allgemeinen kognitiven Fähigkeit weitestgehend identisch (siehe Anhang).

(3) *Konfirmatorische Mehrgruppen-Faktorenanalysen.* Zur Analyse der generellen Differenzierungshypothese (Hypothese 2a) wurden konfirmatorische Mehrgruppen-Faktorenanalysen mit 8, 16 und 32 Fähigkeitsgruppen als Gruppierungsvariable durchgeführt (Abb. 22a, in dieser Abbildung ist nur das Modell für 8 Fähigkeitsgruppen dargestellt).

Die schulformspezifische Differenzierungshypothese konnte nur auf Grundlage von maximal 8 schulformspezifischen Fähigkeitsgruppen geprüft werden (Abb. 22b), da Mplus 3.01 aufgrund einer programmbedingten Beschränkung der Gruppenanzahl keine Differenzierung in 16 oder 32 schulformspezifische Fähigkeitsgruppen zuließ. Die Tatsache, dass einige schulformspezifische Fähigkeitsgruppen von den Analysen ausgeschlossen wurden, ist in Abbildung 22b durch die leeren Felder bei den Fähigkeitsgruppen 7 und 8 bei der Hauptschule und den leeren Felder 1 und 2 für das Gymnasium dargestellt.

Durch die Verwendung der Fähigkeitsgruppen als Gruppierungsvariable wurde die Varianz der allgemeinen kognitiven Fähigkeit (mit Ausnahme der beiden Extremgruppen) weitestgehend konstant gehalten (siehe Anhang). Der Versuch scheiterte, zusätzlich modellbasiert innerhalb der Fähigkeitsgruppen für interindividuelle Unterschiede in der allgemeinen kognitiven Fähigkeit zu kontrollieren: In einem Nested-Faktormodell, in dem nur vier Fähigkeitsgruppen unterschieden wurden, wurden für drei von vier Gruppen negative Faktorvarianzen von allgemeiner kognitiver Fähigkeit geschätzt. Bei einer noch differenzierteren Fähigkeitsgruppierung ist davon auszugehen, dass dieses Problem noch verstärkt auftreten würde (siehe Carlstedt, 2001, S. 597–598, der ein ähnliches Problem beschreibt).

Aus diesem Grund wurde – wie bei Carlstedt – kein vollständiges Nested-Faktormodell spezifiziert. Die Varianz der mathematikspezifischen Fähigkeit wurde auf Grundlage der vier stoffgebietsspezifischen WLE-Scores im Rahmen eines einfaktoriellen Mehrgruppen-Modells mit den Fähigkeitsgruppen als Gruppierungsvariable analysiert (Abb. 22a und 22b).

Zur inferenzstatistischen Absicherung der beiden Differenzierungshypothesen wurden drei Modellvarianten -a, -b und -c analysiert:

- Bei der Modellvariante -a wurden die Varianz der mathematikspezifischen Fähigkeit, die unstandardisierten Faktorladungen sowie die Varianzen der Residualterme der manifesten Variablen über die Gruppen hinweg invariant gesetzt. Zur Identifikation des Faktors  $M'$  wurde die unstandardisierte Faktorladung des Algebra-WLE-Scores in allen Gruppen und allen drei Modellvarianten auf 1,0 fixiert.
- Bei der Modellvariante -b wurden die Varianz der mathematikspezifischen Fähigkeit und die unstandardisierten Faktorladungen über die Gruppen hinweg invariant gesetzt. Die Residualterme konnten gruppenübergreifend variieren.



**(a) Generelle Differenzierungshypothese auf Basis der Analysestichprobe**

**(b) Schulformspezifische Differenzierungshypothese in Abhängigkeit von der Schulform**

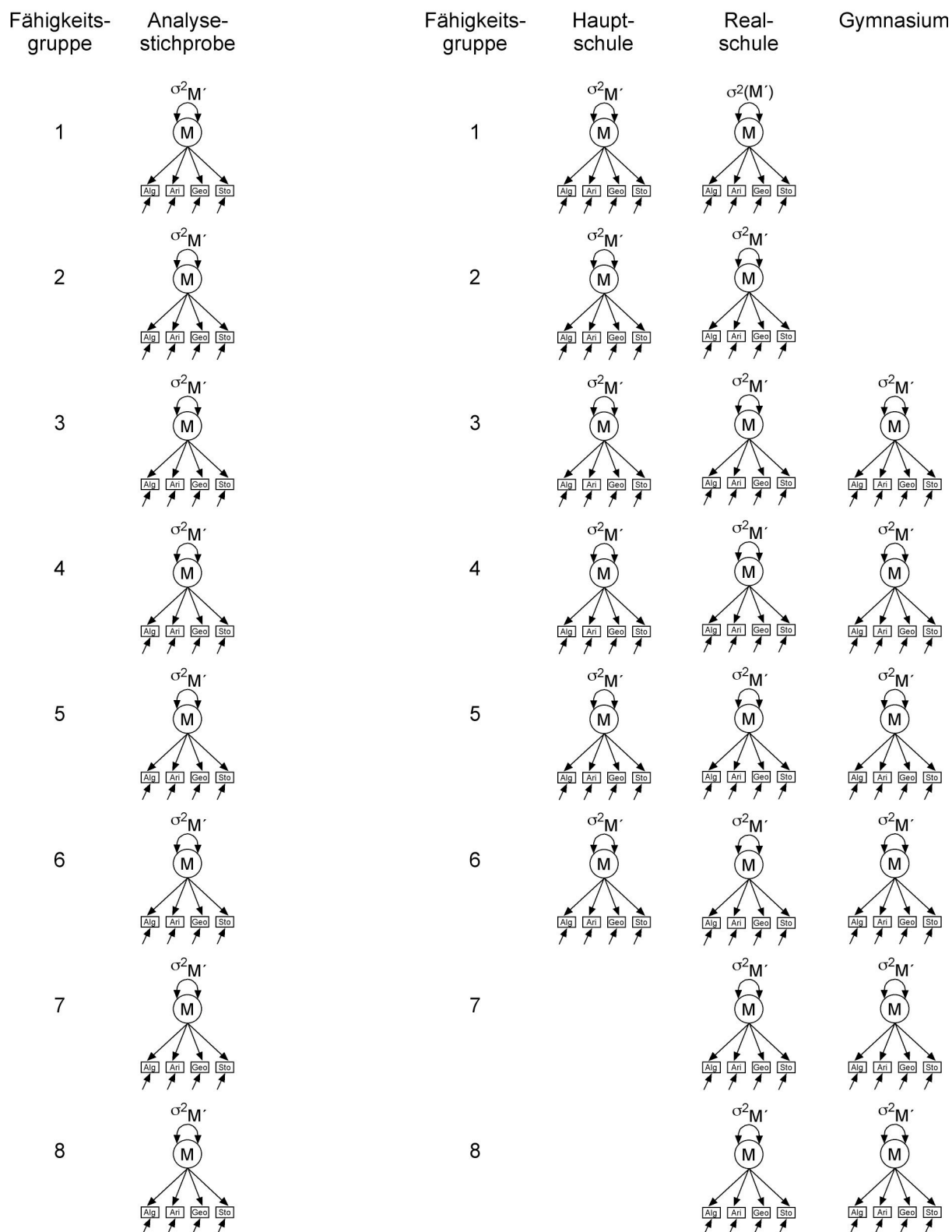


Abbildung 22: Modell zur Prüfung der Differenzierungshypothesen

- Bei der Modellvariante -c wurden nur die unstandardisierten Faktorladungen gruppenübergreifend invariant gesetzt. Residualterme und die Faktorvarianz der mathematikspezifischen Fähigkeit konnten frei variieren. Eine Zunahme der Faktorvarianz der mathematikspezifischen Fähigkeit mit zunehmendem Fähigkeitsniveau indizierte die Differenzierung der mathematikspezifischen Fähigkeit (Carlstedt, 2001).

Wie bei Carlstedt wurde zunächst Variante -a, dann -b und zuletzt -c geprüft. Da die latenten Mittelwerte der mathematikspezifischen Fähigkeit für die Analyse der Differenzierung nicht von Interesse waren, konnten analog zu Carlstedt (2001) die Intercepts der manifesten Variablen über die Gruppen hinweg in allen drei Modellvarianten variieren.

(4) *Simulation von Daten.* Als eine zusätzliche Vergleichsmarke zur Absicherung der generellen Differenzierungshypothese wurde eine Stichprobe von 29.386 „Personen“ mit Mplus 3.01 generiert. In der Population, aus der diese Stichprobe stammte, galt das Nested-Faktormodell. Im Simulationsmodell wurde aber keine generelle Differenzierung der mathematikspezifischen Fähigkeit spezifiziert. Zur Simulation der Daten wurden alle Populationsparameter auf die unstandardisierten Modellparameter fixiert, die auf Grundlage der vollständigen Analysestichprobe von 29.386 Schülern für das Nested-Faktormodell (Modell 7a der ersten Forschungsfrage) geschätzt wurden. Auf Basis der simulierten Daten wurden die drei Analyseschritte nach der Methode von Carlstedt durchgeführt.

### 8.2.3 Ergebnisse

Zunächst werden die Ergebnisse der generellen Differenzierungshypothese dargestellt. Daran schließen die Ergebnisse zur schulformspezifischen Differenzierungshypothese an.

(1) *Generelle Differenzierungshypothese.* Zur Prüfung der generellen Differenzierungshypothese war es zunächst notwendig, den Modell-Fit zu evaluieren. Wenn dieser für die Modellvariante -c akzeptabel war, konnte im nächsten Schritt die Varianz der mathematikspezifischen Fähigkeit in Abhängigkeit vom Fähigkeitsniveau analysiert werden.

*Modellevaluation.* Von den drei Modellvarianten wies – gemessen an den deskriptiven Fit-Indizes – die Modellvariante -c insgesamt den besten Modell-Fit auf (Tab. 21). Dieses Befundmuster zeigte sich für 8, 16 und 32 Fähigkeitsgruppen. Der etwas niedrige Wert des CFI (bei allen drei Fähigkeitsgruppen) konnte dadurch erklärt werden, dass insgesamt gesehen die stoffgebietspezifischen WLE-Scores nur gering interkorrelierten. Aus den niedrigen Interkorrelationen folgte, dass bereits das Baseline-Modell, in dem die manifesten Variablen nicht interkorrelierten, relativ guten Modell-Fit aufwies. Gleichzeitig konnte nur ein geringer Anteil der Varianz durch den Faktor erklärt werden, der die mathematikspezifische Fähigkeit repräsentierte. Die inkrementelle Verbesserung des Modell-Fits durch die Aufhebung der Restriktion fähigkeitsgruppenübergreifender gleicher Varianzen von  $M'$  war damit in Relation zum Baseline-Modell gering. Dies schlug sich in dem CFI-Wert von .89 nieder (siehe auch Brunner & Süß, 2005 für diese Argumentation). Betrachtete man aber zusätzlich zum CFI noch die absolute Abweichung von modell implizierten und empirischen Daten (gemessen durch den SRMR), so war ersichtlich, dass die Modellvariante -c die empirischen Relationen gut approximiert. Diese Schlussfolgerung wurde auch durch den Wert des RMSEA gestützt.

Tabelle 21: Generelle Differenzierungshypothese der mathematikspezifischen Fähigkeit:  
Deskriptive Modell-Fit-Indizes für die Analysestichprobe

Modell	$\chi^2$	df	SF	CFI	RMSEA	SRMR
<i>8 Fähigkeitsgruppen</i>						
Modell -a	1.066	72	1.117	.65	.06	.08
Modell -b	660	44	1.052	.79	.06	.05
Modell -c	357	37	1.021	.89	.05	.03
<i>16 Fähigkeitsgruppen</i>						
Modell -a	912	152	1.109	.66	.05	.07
Modell -b	478	92	1.037	.83	.05	.04
Modell -c	338	77	1.017	.88	.04	.03
<i>32 Fähigkeitsgruppen</i>						
Modell -a	1.001	312	1.107	.66	.05	.07
Modell -b	515	188	1.036	.84	.04	.04
Modell -c	398	157	1.022	.88	.04	.03

Alle  $\chi^2$ -Goodness-of-Fit-Statistiken waren signifikant ( $p < .001$ ). Für eine Beschreibung der Modellspezifikation siehe Abschnitt 8.2.2.

Tabelle 22: Generelle Differenzierungshypothese: Deskriptive Modell-Fit-Indizes für die  
simulierten Daten

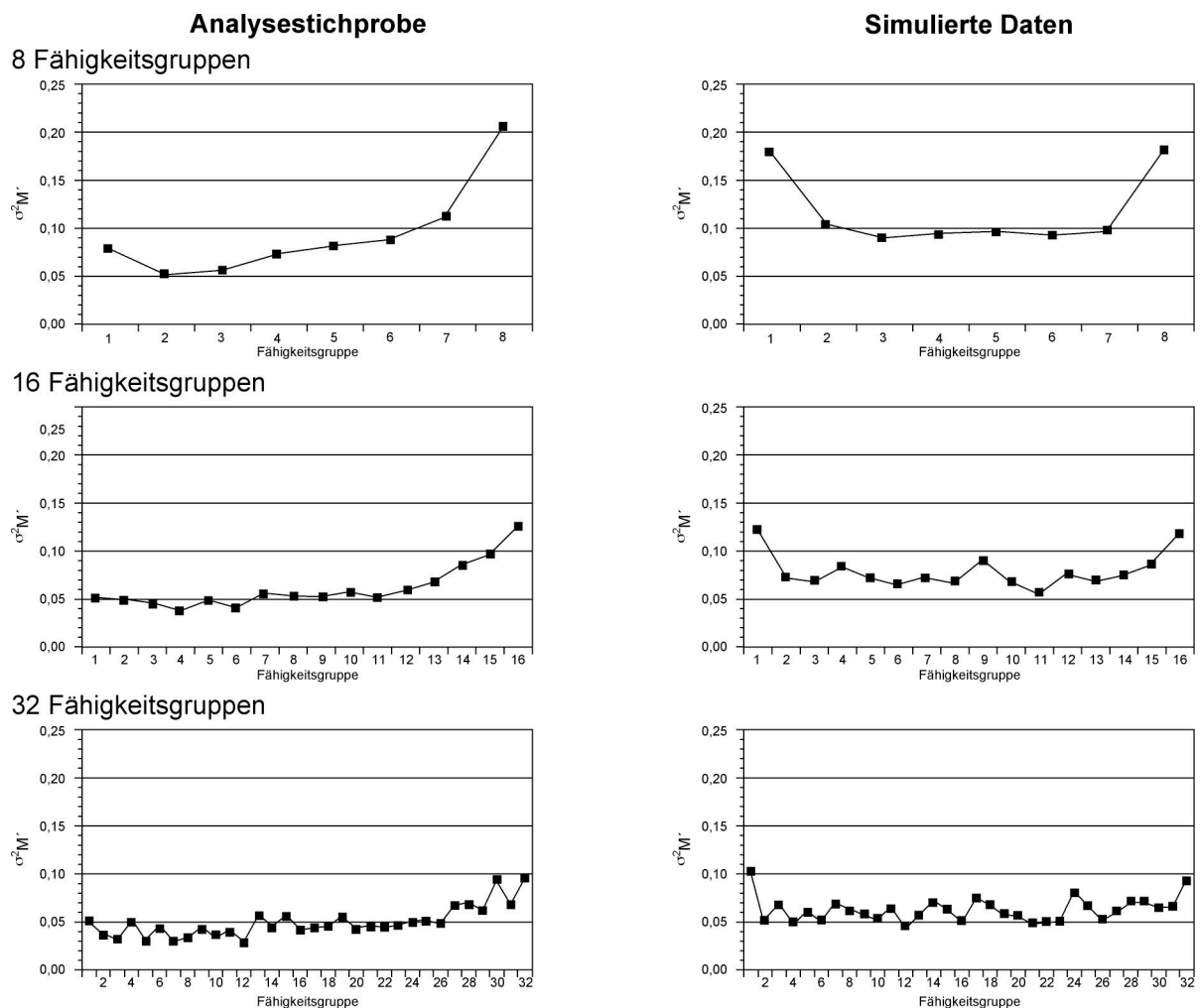
Modell	$\chi^2$	df	CFI	RMSEA	SRMR
<i>8 Fähigkeitsgruppen</i>					
Modell -a	404	72	.89	.04	.05
Modell -b	324	44	.91	.04	.04
Modell -c	99	37	.98	.02	.02
<i>16 Fähigkeitsgruppen</i>					
Modell -a	278	152	.95	.02	.03
Modell -b	198	92	.95	.03	.03
Modell -c	121	77	.98	.02	.02
<i>32 Fähigkeitsgruppen</i>					
Modell -a	406	312	.95	.02	.04
Modell -b	249	188	.97	.02	.03
Modell -c	186	157	.99	.01	.02

Für eine Beschreibung der Modellspezifikation siehe Abschnitt 8.2.2.

*Modellevaluation der simulierten Daten.* Auch bei den simulierten Daten stützten die Werte des CFI, des RMSEA und des SRMR Modellvariante -c, die letztlich als beste Approximation betrachtet wurde (Tab. 22).

Die Evaluation der realen und der simulierten Daten zeigte also, dass jeweils das Modell -c, in dem die Varianzen der mathematikspezifischen Fähigkeit über die Fähigkeitsgruppen hinweg variieren konnten, die realen oder simulierten Daten am besten approximiert. Doch stützten die Ergebnisse auch die generelle Differenzierungshypothese?

*Generelle Differenzierung der mathematikspezifischen Fähigkeit.* Die deskriptive Zunahme der Faktorvarianz der mathematikspezifischen Fähigkeit mit zunehmendem Fähigkeitsniveau in Abbildung 23 zeigte, dass die Differenzierung der mathematikspezifischen Fähigkeit in der Analysestichprobe vorzufinden war.<sup>27</sup> Dieser Befund konnte auch anhand einer Stichprobe repliziert werden, in der sich nur die Daten der Schüler befanden, die zur Prüfung der schulformspezifischen Differenzierungshypothese verwendet wurden (siehe Abb. 25).



Vergleich der Differenzierung in der Analysestichprobe und den korrespondierenden simulierten Daten.

Abbildung 23: Generelle Differenzierungshypothese der mathematikspezifischen Fähigkeit ( $M'$ )

<sup>27</sup> Die Zunahme der Varianz von  $M'$  in den beiden Extremgruppen (bei den realen und den simulierten Daten) resultierte daher, weil in diesen Gruppen die Varianz der allgemeinen kognitiven Fähigkeit nicht vollständig konstant gehalten werden konnte (siehe Carlstedt, 2001, der mit dem gleichen Problem konfrontiert war).

Der Graph für die Stichprobe mit den simulierten Daten (Abb. 23, Spalte „simulierte Daten“), zeigte, dass – wie zu erwarten – keine Differenzierung der mathematikspezifischen Fähigkeit vorzufinden war, da diese nicht explizit im Simulationsmodell spezifiziert wurde. Trotz identischer Modellparameter war also die Differenzierung der mathematikspezifischen Fähigkeit mit zunehmendem Fähigkeitsniveau nur in den realen Daten vorzufinden.

Wichtig bei der Analyse von Abbildung 23 ist aber zu beachten, dass die Faktorvarianzen *nicht* fähigkeitsgruppenübergreifend oder zwischen den realen und simulierten Daten verglichen werden konnten. Der Grund hierfür war, dass die unstandardisierten Faktorladungen sich zwischen den Modellen für die unterschiedliche Fähigkeitsgruppenanzahl bei den realen sowie den simulierten Daten unterschieden. Somit waren die Faktoren, die die mathematikspezifische Fähigkeit repräsentieren, fähigkeitsgruppen- und stichprobenübergreifend nicht metrisch invariant.

Die deskriptive Zunahme der Faktorvarianz von M' konnte auch inferenzstatistisch durch einen Vergleich der  $\chi^2$ -Differenzen abgesichert werden (Tab. 23): Dies indizierte die signifikante Verbesserung des  $\chi^2$ -*Goodness-of-Fit*-Tests beim Übergang von Modellvariante -b zu Modellvariante -c. Die generelle Differenzierungshypothese (Hypothese 2a) wurde somit deskriptiv und inferenzstatistisch gestützt.

Tabelle 23: Inferenzstatistische Prüfung der generellen Differenzierungshypothese

Modellvergleich	$\Delta\chi^2$	$\Delta df$
<i>8 Fähigkeitsgruppen</i>		
Modell -a versus Modell -b	407*	28
Modell -b versus Modell -c	272*	7
<i>16 Fähigkeitsgruppen</i>		
Modell -a versus Modell -b	423*	60
Modell -b versus Modell -c	134*	15
<i>32 Fähigkeitsgruppen</i>		
Modell -a versus Modell -b	473*	124
Modell -b versus Modell -c	115*	31

\*  $p < .05$ .

Zur Berechnung der  $\chi^2$ -Differenz wurde die Korrekturformel aus dem technischen Anhang von Mplus verwendet (Muthén & Muthén, 1998–2004a, S. 22).

(2) *Schulformspezifische Differenzierungshypothese*. Wie auch bei der Analyse der generellen Differenzierungshypothese, wurde bei Prüfung der schulformspezifischen Differenzierungshypothese zunächst der Modell-Fit der spezifizierten Modelle evaluiert.

*Modellevaluation*. Von den drei Modellvarianten wies gemessen an den deskriptiven Fit-Indizes die Modellvariante -c den besten Modell-Fit auf (Tab. 24).

Tabelle 24: Schulformspezifische Differenzierungshypothese: Deskriptive Modell-Fit-Indizes

Modell	$\chi^2$	df	SF	CFI	RMSEA	SRMR
Modell -a	1.145	192	1.120	.57	.07	.10
Modell -b	706	116	1.036	.74	.07	.06
Modell -c	396	97	1.007	.87	.06	.04

Für eine Beschreibung der Modellspezifikation siehe Abschnitt 8.2.2. SF = Skalierungsfaktor.

Der etwas niedrige Wert des CFI konnte – wie auch bei den Analysen zur generellen Differenzierungshypothese (siehe oben) – dadurch erklärt werden, dass insgesamt gesehen die stoffgebietspezifischen WLE-Scores nur gering interkorrelierten. Dies schlug sich in dem CFI-Wert von .87 nieder. Betrachtete man aber zusätzlich zum CFI noch die absolute Abweichung von modellimplizierten und empirischen Daten (gemessen durch den SRMR), so war ersichtlich, dass die Modellvariante -c die empirischen Relationen gut approximiert. Diese Schlussfolgerung wurde auch durch den Wert des RMSEA gestützt. Da sich der Modell-Fit gemessen durch den CFI, RMSEA und SRMR jeweils von der Modellvariante -a über -b hin zu -c verbesserte (oder zumindest gleich blieb), wurde letztlich Modellvariante -c favorisiert.

Diese Verbesserungen des Modell-Fits waren auch statistisch signifikant. So führte das Aufheben der Restriktion gleicher Residualvarianzen (Modell -a vs. -b) zu einer signifikanten Verbesserung des Modell-Fits ( $\Delta\chi^2 = 329$ ,  $df = 28$ ,  $p < .001$ ). Der inferenzstatistische Vergleich der Modelle -b und -c zeigte auch, dass die Varianzen von  $M'$  in den schulformspezifischen Fähigkeitsgruppen nicht gleich groß waren ( $\Delta\chi^2 = 239$ ,  $df = 7$ ,  $p < .001$ ). Doch war die Differenzierung im Sinne von Hypothese 2b von der Schulform abhängig?

*Schulformspezifische Differenzierung der mathematikspezifischen Fähigkeit.* Abbildung 24 stellt die Ergebnisse zur schulformspezifischen Differenzierung dar. Hierbei ist zu beachten, dass die Abnahme der Varianz von  $M'$  von Fähigkeitsgruppe 1 auf 2 für Realschüler und Hauptschüler sowie die Zunahme der Differenzierung von Fähigkeitsgruppe 7 auf 8 für Realschüler und Gymnasiasten nur eingeschränkt interpretiert werden können, da in die Varianz der mathematikspezifischen Fähigkeit für diese Schülergruppen auch Varianz der allgemeinen kognitiven Fähigkeit einging (vgl. Anhang).

Als globaler inferenzstatistischer Test, ob die Differenzierung der mathematikspezifischen Fähigkeit von der Schulform abhängig war, wurde Modell -c mit einem Modell verglichen, in dem die Varianz von  $M'$  für Schüler gleicher Fähigkeitsgruppen schulformübergreifend auf den gleichen Wert fixiert wurde. Die resultierende  $\chi^2$ -Differenz ( $\Delta\chi^2 = 27$ ,  $df = 12$ ,  $p < .01$ ) war statistisch signifikant. Damit konnte auf einer allgemeinen Ebene geschlossen werden, dass Schüler gleicher Fähigkeitsgruppen, aber unterschiedlicher Schulformen nicht die gleiche Differenzierung der mathematikspezifischen Fähigkeit aufwiesen.

Jedoch war – wie die deskriptive Analyse von Abbildung 24 zeigte – die Differenzierung nur für Schüler im jeweils unteren schulformspezifischen Fähigkeitspektrum unterschiedlich: Die Heterogenität der mathematikspezifischen Fähigkeit von Gymnasiasten war im unteren Fähigkeitspektrum (Fähigkeitsgruppen 3 und 4) tendenziell größer als bei Schüler der anderen beiden Schulformen. Weiterhin war die mathematikspezifische Varianz von Realschülern in den Fähigkeitsgruppen 1 und 2 größer als bei Hauptschülern. Jedoch verschwanden die Unterschiede in der

Differenzierung zwischen der Realschule und der Hauptschule nahezu in den Fähigkeitsgruppen 3 und 4, bzw. Schüler aller drei Schulformen wiesen für Fähigkeitsgruppe 5 und 7 nahezu die gleiche Differenzierung der mathematikspezifischen Fähigkeit auf. Erwartungswidrig kehrte sich die Differenzierung in Fähigkeitsgruppe 6 sogar um: Hier war die Differenzierung bei Hauptschülern etwas stärker als bei Realschülern und Letztere wiederum war etwas stärker als bei Gymnasiasten.

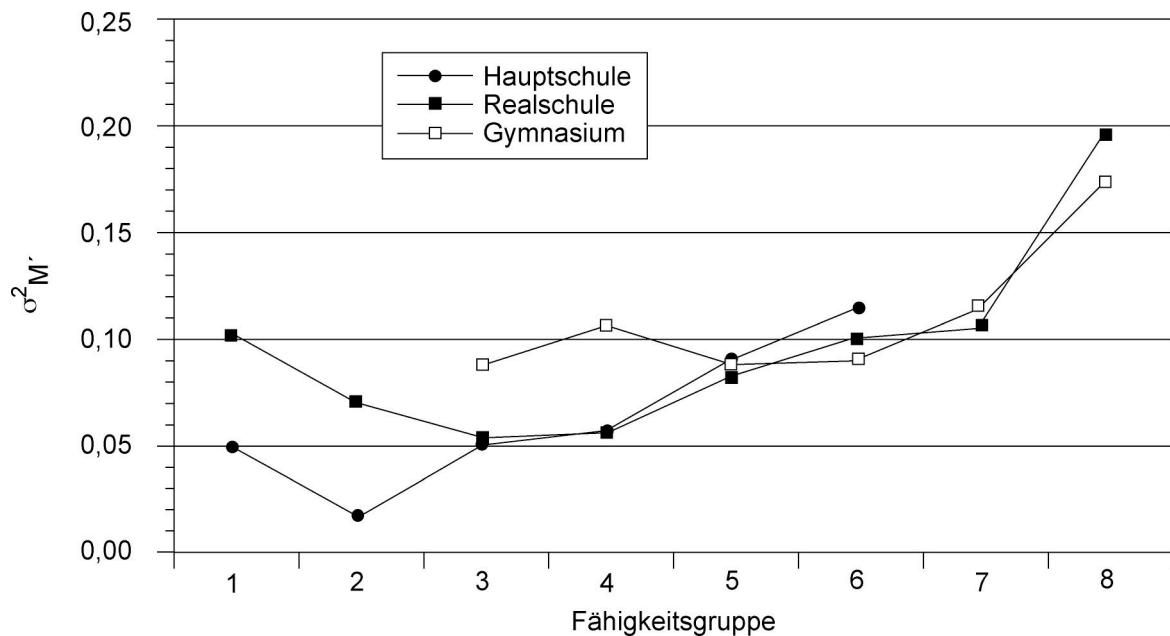


Abbildung 24: Differenzierung der mathematikspezifischen Fähigkeit in Abhängigkeit des Fähigkeitsniveaus und der Schulformzugehörigkeit

Zur inferenzstatistischen Prüfung der unterschiedlichen schulformspezifischen Differenzierung (innerhalb einer Fähigkeitsgruppe) wurde Modell -c mit einem Modell verglichen, in dem ein Varianzpaar von zwei Schulformen innerhalb derselben Fähigkeitsgruppe auf den gleichen Wert fixiert wurde. Die Faktorvarianzen von  $M'$  in den restlichen Fähigkeitsgruppen wurden frei geschätzt. Eine signifikante  $\chi^2$ -Differenz indizierte eine ungleiche schulformspezifische Differenzierung der mathematikspezifischen Fähigkeit innerhalb einer Fähigkeitsgruppe.

Von den insgesamt 16 Paarvergleichen waren nur drei statistisch signifikant (Tab. 25). Dies waren die Vergleiche der Faktorvarianzen von  $M'$  zwischen Haupt- und Realschülern in den Fähigkeitsgruppen 1 und 2 sowie der Vergleich von Realschülern und Gymnasiasten in der Fähigkeitsgruppe 4. Für alle anderen Vergleiche der Faktorvarianzen zwischen den Schulformen konnte *nicht* zufallskritisch geschlussfolgert werden, dass sich die Differenzierung der mathematikspezifischen Fähigkeit signifikant unterschieden hatte.

Tabelle 25: Inferenzstatistische Prüfung der schulformspezifischen Differenzierungshypothese

Vergleich	$\chi^2$	SF	$\Delta\chi^2$	$\Delta df$
<i>Fähigkeitsgruppe 1</i>				
$\sigma^2_{M'}(\text{HS}) = \sigma^2_{M'}(\text{RS})$	403	1,010	5,66*	1
<i>Fähigkeitsgruppe 2</i>				
$\sigma^2_{M'}(\text{HS}) = \sigma^2_{M'}(\text{RS})$	407	1,009	9,32*	1
<i>Fähigkeitsgruppe 3</i>				
$\sigma^2_{M'}(\text{HS}) = \sigma^2_{M'}(\text{RS})$	396	1,009	0,16	1
$\sigma^2_{M'}(\text{HS}) = \sigma^2_{M'}(\text{GY})$	396	1,011	1,04	1
$\sigma^2_{M'}(\text{RS}) = \sigma^2_{M'}(\text{GY})$	396	1,011	0,76	1
<i>Fähigkeitsgruppe 4</i>				
$\sigma^2_{M'}(\text{HS}) = \sigma^2_{M'}(\text{RS})$	396	1,008	0,10	1
$\sigma^2_{M'}(\text{HS}) = \sigma^2_{M'}(\text{GY})$	400	1,009	3,39	1
$\sigma^2_{M'}(\text{RS}) = \sigma^2_{M'}(\text{GY})$	401	1,011	4,57*	1
<i>Fähigkeitsgruppe 5</i>				
$\sigma^2_{M'}(\text{HS}) = \sigma^2_{M'}(\text{RS})$	397	1,006	0,37	1
$\sigma^2_{M'}(\text{HS}) = \sigma^2_{M'}(\text{GY})$	397	1,006	0,08	1
$\sigma^2_{M'}(\text{RS}) = \sigma^2_{M'}(\text{GY})$	396	1,008	0,29	1
<i>Fähigkeitsgruppe 6</i>				
$\sigma^2_{M'}(\text{HS}) = \sigma^2_{M'}(\text{RS})$	397	1,005	0,11	1
$\sigma^2_{M'}(\text{HS}) = \sigma^2_{M'}(\text{GY})$	398	1,005	0,95	1
$\sigma^2_{M'}(\text{RS}) = \sigma^2_{M'}(\text{GY})$	397	1,007	0,93	1
<i>Fähigkeitsgruppe 7</i>				
$\sigma^2_{M'}(\text{RS}) = \sigma^2_{M'}(\text{GY})$	397	1,007	0,47	1
<i>Fähigkeitsgruppe 8</i>				
$\sigma^2_{M'}(\text{RS}) = \sigma^2_{M'}(\text{GY})$	398	1,006	0,75	1

\*  $p < .05$ .

Zur Berechnung der  $\chi^2$ -Differenz wurde die Korrekturformel aus dem technischen Anhang von Mplus verwendet (Muthén & Muthén, 1998–2004a, S. 22). SF = Skalierungsfaktor.



## 8.2.4 Diskussion

Die wichtigsten Ergebnisse der Analysen der beiden Differenzierungshypothesen können wie folgt zusammengefasst werden:

- Die generelle Differenzierungshypothese der mathematikspezifischen Fähigkeit wurde von den Ergebnissen gestützt. Das bedeutet, dass die Heterogenität der mathematikspezifischen Fähigkeit in der Gesamtstichprobe mit dem Niveau allgemeiner kognitiver Fähigkeit zunahm.
- Die schulformspezifische Differenzierungshypothese der mathematikspezifischen Fähigkeit wurde nur teilweise deskriptiv wie auch inferenzstatistisch von den Daten gestützt. Nur für einige wenige Schülergruppen war bei gleichem Niveau allgemeiner kognitiver Fähigkeit die Heterogenität der mathematikspezifischen Fähigkeit an höheren Schulformen stärker.

Nachfolgend werden zwei Fragen zu diesen Ergebnissen erörtert:

- (1) Wie können die Ergebnisse zur schulformspezifischen Differenzierung erklärt werden?
- (2) Wie können die Ergebnisse zur schulformspezifischen und generellen Differenzierung in Einklang gebracht werden?

(1) *Schulformspezifische Differenzierungshypothese.* Die schulformspezifische Differenzierungshypothese wurde in Anlehnung an Abad und Kollegen (2003) dadurch erklärt, dass Schüler mit höherem Fähigkeitsniveau in der Regel höhere Schulformen besuchen und dort bessere Investitionsmöglichkeiten für ihre allgemeine kognitive Fähigkeit vorfinden (siehe auch Jensen, 1998). Wenn Schüler das gleiche Niveau allgemeiner kognitiver Fähigkeit erreichen, sollten die Schüler an höheren Schulformen eine stärkere Differenzierung der mathematikspezifischen Fähigkeit aufweisen.

Angesichts der Ergebnisse der vorliegenden Arbeit scheint diese Erklärung nur für die jeweils leistungsschwächeren Schüler (indiziert durch die schulformspezifische Fähigkeitsgruppe) der jeweiligen Schulform zu greifen: Realschüler in den Fähigkeitsgruppen 1 und 2 wiesen eine stärkere Differenzierung auf als Hauptschüler dieser Fähigkeitsgruppen. Gymnasiasten der Fähigkeitsgruppen 3 und 4 wiesen zumindest deskriptiv eine stärkere Differenzierung auf als Realschüler und Hauptschüler. Man kann dieses Befundmuster so interpretieren, dass schulformspezifische Lernangebote insbesondere die schwächeren Schüler an Realschulen und Gymnasien zur Investition ihrer allgemeinen kognitiven Fähigkeit anregen: Schüler dieser Schulformen nutzen diese Lernangebote differenziell in Abhängigkeit von ihrem Niveau allgemeiner kognitiver Fähigkeit.

Ab einschließlich Fähigkeitsgruppe 3 nahm mit zunehmendem Fähigkeitsniveau bei Haupt- und Realschülern sowie ab Fähigkeitsgruppe 6 (wenn auch nur schwach) bei Gymnasiasten die Varianz der mathematikspezifischen Fähigkeit tendenziell zu. Weiterhin war die Differenzierung in diesen Fähigkeitsgruppen schulformübergreifend nahezu in gleicher Weise ausgeprägt. Möglicherweise greift in allen diesen Schulformen die Erklärung, die auch für die generelle Differenzierungshypothese der mathematikspezifischen Fähigkeit gegeben wurde: Mit Jensen (1998) kann man die Differenzierung dadurch erklären, dass Schüler unterschiedlichen Fähigkeitsniveaus (unabhängig von der Schulform) sich darin unterscheiden, inwiefern sie problemlöserrelevantes Mathematikwissen prozeduralisiert haben. Dieser Mechanismus ersetzt oder überlagert möglicherweise den Mechanismus unterschiedlicher Investitionsmöglichkeiten in Abhängigkeit von der Schulform.

(2) *Schulformspezifische und generelle Differenzierung.* Wie können die Ergebnisse zur schulformspezifischen und generellen Differenzierung in Einklang gebracht werden? Ergebnis der Analysen zur generellen Differenzierungshypothese war, dass mit zunehmendem Fähigkeitsniveau die Differenzierung der mathematikspezifischen Fähigkeit zunahm. Betrachtet man hingegen Abbildung 24, so ist diese generelle Differenzierung nicht unmittelbar ersichtlich. Hierbei ist jedoch zu bedenken, dass die schulformspezifischen Fähigkeitsgruppen nicht mit der gleichen Schülerzahl besetzt waren (vgl. Abb. 21). Weiterhin nahm die Differenzierung der mathematikspezifischen Fähigkeit schulformübergreifend in höheren Fähigkeitsgruppen zu. Bei den Analysen zur generellen Differenzierungshypothese überlagerte sich die Differenzierung von  $M'$  der verschiedenen Schulformen: Die Schüler gingen gewichtet entsprechend der schulformspezifischen Fähigkeitsgruppengröße in die Berechnung der Varianz von  $M'$  ein. Diese Gewichtung führte zur generellen Differenzierung.

Abbildung 25 veranschaulicht dies: In der Abbildung ist zum einen die generelle Differenzierung für 8 Fähigkeitsgruppen eingezeichnet, wie sie für die Stichprobe zur Analyse der schulformspezifischen Differenzierung bestimmt wurde (die Schulformzugehörigkeit wurde also bei den Analysen nicht berücksichtigt). Zum anderen ist eine *geschätzte* Differenzierung von  $M'$  für diese Stichprobe als gestrichelte Linie eingezeichnet, in der die schulformspezifische Differenzierung von  $M'$  mit der schulformspezifischen Fähigkeitsgruppengröße gewichtet wurde. Man sieht, dass die Kurven nahezu identisch sind. Allerdings ist bei Abbildung 25 zu beachten, dass es *nicht* möglich ist, die Absolutwerte der Faktorvarianzen von  $M'$  der geschätzten Differenzierung mit der tatsächlichen Differenzierung zu vergleichen. Der Grund hierfür ist, dass die Metrik der mathematikspezifischen Fähigkeit zwischen den beiden Modellen nicht vergleichbar war, da sich die korrespondierenden (unstandardisierten) Faktorladungen zwischen den beiden Modellen unterscheiden. Dennoch zeigt der nahezu parallele Verlauf beider Linien den Effekt der Gewichtung der schulformspezifischen Differenzierung der mathematikspezifischen Fähigkeit recht deutlich.

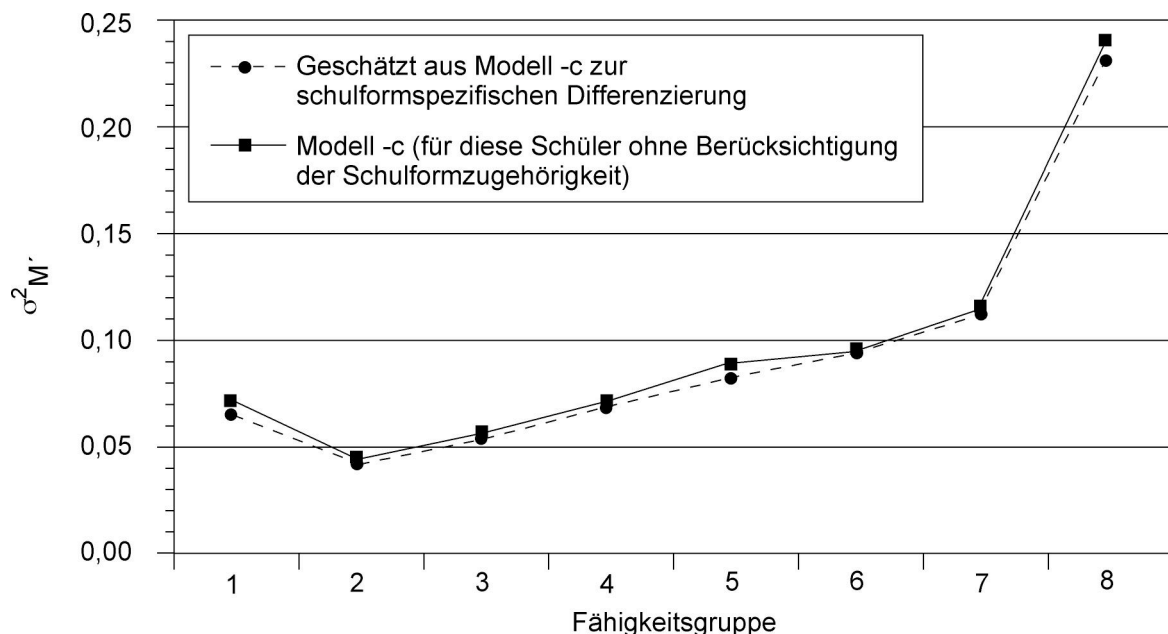


Abbildung 25: Gegenüberstellung der aus dem Modell zur schulformspezifischen Differenzierung der mathematikspezifischen Fähigkeit geschätzten und der tatsächlichen Differenzierung in dieser Stichprobe

Die Annahme, dass sich Unterschiede in der schulformspezifischen Differenzierung bei der generellen Differenzierung überlagern, wurde auch durch die Ergebnisse zum Nested-Faktormodell im Rahmen der Mehrgruppen-Faktorenanalyse in Abschnitt 8.1 gestützt. Es zeigte sich, dass die „mittlere“ Heterogenität der Schüler hinsichtlich ihrer mathematikspezifischen Fähigkeit von der Hauptschule über die Realschule hin zum Gymnasium zunahm. Nun ist zu bedenken, dass der Besuch höherer Schulformen mit einem höheren Niveau allgemeiner kognitiver Fähigkeit einherging (vgl. Abb. 19). Daraus folgte, dass – wenn die Schulform nicht berücksichtigt wurde – mit zunehmendem Fähigkeitsniveau die Differenzierung der mathematikspezifischen Fähigkeit zunahm.

Insgesamt gesehen kann spekuliert werden, ob die generelle Differenzierung der mathematikspezifischen Fähigkeit durch die Überlagerung von zwei Mechanismen zu Stande kommt:

- Erstens können Schüler durch schulformspezifische Unterschiede in den Investitionsmöglichkeiten (für ihre allgemeine kognitive Fähigkeit) zu differenziellem Investitionsverhalten angeregt werden. Dieser Mechanismus greift bei den Schülern der jeweiligen Schulformen, die schwächere Leistungen bei der allgemeinen kognitiven Fähigkeit erzielen.
- Zweitens nutzen Schüler Lernangebote im Mathematikunterricht (unabhängig vom schulformspezifischen Niveau dieser Lernangebote) differenziell. Dieser Mechanismus greift bei den Schülern der jeweiligen Schulformen, die stärkere Leistungen bei der allgemeinen kognitiven Fähigkeit erzielen.

Die Vermutung einer Überlagerung beider Mechanismen ist wie gesagt spekulativ und bedarf weiterer Forschung und natürlich der Replikation. Einige methodische Überlegungen werden hierzu im Rahmen der Gesamtdiskussion erörtert.

### 8.3 Zusammenfassende Diskussion zu Schulformunterschieden

Die zweite Forschungsfrage beschäftigte sich in erster Linie mit Schulformunterschieden bei den kognitiven Fähigkeiten mathematischer Schülerleistung. Im Rahmen der Gesamtdiskussion soll nun die Frage erörtert werden, welche methodischen Anforderungen idealerweise erfüllt sein sollten, um Schulformunterschiede zu analysieren.

In der vorliegenden Arbeit wurden Mittelwertunterschiede der kognitiven Fähigkeiten der mathematischen Schülerleistung sowie Unterschiede in der Heterogenität der mathematikspezifischen Fähigkeit zwischen den Schulformen untersucht. Für diese Analysen und eine eindeutige Interpretation der Ergebnisse müssen einige zentrale Voraussetzungen erfüllt sein. Diese werden nachfolgend weiter erörtert.

*Messinvarianz.* Die wichtigste Voraussetzung für die Analyse von Schulformunterschieden ist die Invarianzprüfung der verwendeten Strukturmodelle. Erst diese ermöglicht es, Schulformvergleiche auf eine methodisch solide Basis zu stellen. In der bisherigen Forschung zu Schulformvergleichen ist bisher nicht in Frage gestellt worden, ob schulformübergreifend das gleiche Strukturmodell die empirischen Relationen ausreichend gut approximiert. Dies trifft für die methodisch sehr fundierten Analysen aus BIJU, TIMSS und PISA genauso zu wie für Analysen aus den Längsschnittuntersuchungen SCHOLASTIK (z.B. Schneider & Stefanek, 2004) oder LAU (Lehmann, Hunger, Ivanov & Gänsfuß, 2004). Wenn Unterschiede zwischen den Schulformen

oder Bildungsgängen untersucht wurden (z.B. Lind & Knoche, 2004), wurden diese im Rahmen von differenziellen Itemfunktionen analysiert (DIF, für einen Überblick siehe z.B. Glöckner-Rist & Hoijtink, 2003; Raju, Laffitte & Byrne, 2002). Bei Verwendung eines eindimensionalen Raschmodells wie in der Arbeit von Knoche und Lind wird aber per se davon ausgegangen, dass alle Items (schulformübergreifend) in gleicher Weise von einer latenten Variablen beeinflusst werden. Die substanzielle Frage (siehe hierzu die Ableitung der Invarianzbedingungen in Abschnitt 7.1.2), ob tatsächlich die Struktur (im Sinne von Faktorladungen oder Itemdiskriminationsparametern) als schulformübergreifend invariant angenommen werden kann, wurde nicht untersucht.

Dies ist auch eine Einschränkung, die auf die Analysen in dieser Arbeit zutrifft: Alle WLE-Scores wurden unter der Annahme schulformübergreifend geltender IRT-Modelle (Raschmodelle für den PISA-Mathematiktest sowie Birnbaummodelle für die Skalen zur Messung der fluiden Fähigkeit anhand der Aufgaben aus dem KFT) geschätzt oder aus den bisherigen Auswertungen von PISA übernommen (PISA-Lesetest). Trotz methodischer Bedenken sprachen vor allem zwei Gründe dafür:

- Erstens war es nur mit dem Programm ConQuest (Wu, Adams & Wilson, 1998) möglich, stabile Schätzungen für die Itemschwierigkeiten zu erhalten. Der Versuch, die Daten der Gesamtstichprobe von 29.386 Schülern für den PISA-Mathematiktest mit dem IRT-Programm Parscale (Muraki & Bock, 1996) zu analysieren, scheiterte wohl am komplexen Multi-Matrix-Sampling-Design und der damit verbundenen großen Anzahl an zufällig erzeugten fehlenden Werten. Da Parscale eines der wenigen Programme ist, mit dem DIF-Analysen für zwei- oder dreiparametrische Birnbaummodelle mit polytomen Daten möglich sind, konnten Invarianzannahmen für den PISA-Mathematiktest nicht auf Itemebene geprüft werden.
- Zweitens konzentrierte sich die zweite Forschungsfrage auf „breite“ kognitive Fähigkeiten und setzte daher auf der aggregierten Ebene von WLE-Scores an. Schulformspezifische Unterschiede der Itemschwierigkeiten oder Itemdiskriminationsparameter waren nicht von Interesse. Daher wurden die Invarianzannahmen der beiden Modelle auch auf Ebene der WLE-Scores analysiert.

Die Frage, ob die Niveauunterschiede zwischen den Schulformen für alle Aufgaben des PISA-Mathematiktests bestehen oder nur für einige wenige, ist empirisch gesehen noch offen. Ein zentraler Leitgedanke bei zukünftigen DIF-Analysen sollte sein, die erwarteten Unterschiede in den Itemschwierigkeiten und Itemdiskriminationsparametern a priori aufgrund theoretischer Annahmen zu formulieren (z.B. auf Grundlage der Annahme unterschiedlicher Unterrichtskulturen, siehe Klieme & Baumert, 2001) und konfirmatorisch zu prüfen (für eine elegante Methode siehe z.B. Glöckner-Rist & Hoijtink, 2003).

*Längsschnittliche Studien zur Unterrichtsforschung.* In dieser Arbeit wurde ein querschnittlicher Analyseansatz zur Analyse der Differenzierungshypothesen verwendet. In der Diskussion der Ergebnisse in Abschnitt 8.2.4 wurde spekuliert, ob die generelle Differenzierung aufgrund der Überlagerung von zwei „Mechanismen“ zu Stande kommen kann.

Der erste Mechanismus ging von schulformspezifischen Niveauunterschieden der Investitionsmöglichkeiten im Mathematikunterricht aus, die die Schüler zur differenziellen Investition ihrer allgemeinen kognitiven Fähigkeit anregen. Inwiefern dieser Mechanismus greift, sollte im Rahmen von längsschnittlichen Studien analysiert werden, in denen kognitive und motivationale Schülermerkmale sowie Indikatoren des Mathematikunterrichts erhoben werden. Um leistungs-

förderliche Aspekte des Mathematikunterrichts zu quantifizieren, ist es optimal, diese Aspekte aus verschiedenen Beobachtungsperspektiven zu erfassen, um ein möglichst umfassendes Bild des Unterrichts zu gewinnen und die differenzielle Validität der verschiedenen Perspektiven für unterschiedliche Zielkriterien des Unterrichts zu berücksichtigen (Clausen, 2002).

Für die statistische Auswertung bieten sich dann Mehrebenen-Modelle an (siehe hierzu z.B. Snijders & Bosker, 1999), wie sie beispielsweise mit dem Programm Mplus (Muthén & Muthén, 1998–2004b) spezifiziert werden können. Wenn Niveauunterschiede der Lernangebote im Mathematikunterricht zur differenziellen Nutzung führen, dann sollte (im Kontext von *Random-Slope*-Modellen) das Gewicht der allgemeinen kognitiven Fähigkeit zur Vorhersage der mathematischen Fähigkeiten mit einer Steigerung des Niveaus lernförderlicher Aspekte des Mathematikunterrichts ebenfalls zunehmen.

Der zweite Mechanismus zur Erklärung der Differenzierung der mathematikspezifischen Fähigkeit ist eine differenzielle Nutzung der Lernangebote im Mathematikunterricht (unabhängig vom Niveau dieser Lernangebote), die zu einer verstärkten Prozeduralisierung problemlöserrelevanten Mathematikwissens führen kann. Indikativ für die differenzielle Nutzung der Lernangebote sind unterschiedliche Regressionsgewichte der Unterrichtsmerkmale zur Vorhersage des Lernzugewinns bei mathematischen Fähigkeiten in Abhängigkeit des Fähigkeitsniveaus der allgemeinen kognitiven Fähigkeit. Als Auswertungsmethode bieten sich Mehrebenen-Mehrgruppenmodelle an, die mit Mplus (Muthén & Muthén, 1998–2004b) analysiert werden können.

Für die Prüfung beider Mechanismen sind zum Beispiel die Daten aus der Längsschnittstudie PISA-I-Plus (Prenzel, Drechsel, Carstensen & Ramm, 2004) und dem daran angedockten COACTIV-Projekt (Krauss u.a., 2004) gut geeignet.

Vor dem Hintergrund der diskutierten methodischen Gesichtspunkte ist klar, dass die Analysen in der vorliegenden Arbeit insbesondere den letzten Punkt „Längsschnittstudien zur Unterrichtsforschung“ nicht erfüllen. Dennoch können diese Analysen die Lehr-Lern-Forschung stimulieren. Insbesondere die möglichen Auswirkungen auf die Heterogenität der Schülerschaft aufgrund einer differenziellen Nutzung der Lernangebote (wie auch die Prüfung von Invarianzannahmen der verwendeten Strukturmodelle) blieben in den bisher methodisch besten Studien (Clausen, 2002; Gruehn, 2000; Kunter, 2005) ausgespart. Im Vordergrund stand der durchschnittliche Effekt von Unterrichtsmerkmalen auf kognitive oder motivationale Zielkriterien des Unterrichts. Die Abhängigkeit der Wirkung des Unterrichts von (motivationalen oder kognitiven) Schülermerkmalen wurde in diesen Studien *nicht* untersucht. Die Analysen in dieser Arbeit machen unter anderem auf diese Forschungslücke aufmerksam, denn die Annahme der differenziellen Nutzung ist zentral im Rahmen des Angebot-Nutzung-Modells, das in der aktuellen Lehr-Lern-Forschung dominiert (Baumert & Köller, 2000; Baumert u.a., 2004; Helmke, 2003; Schatz Koehler & Grouws, 1992).

Über den Beitrag zur Lehr-Lern-Forschung hinaus konnte mit den Analysen im Rahmen der zweiten Forschungsfrage ein sehr differenziertes und methodisch fundiertes Bild über Schulformunterschiede hinsichtlich Niveau und Heterogenität der kognitiven Fähigkeiten mathematischer Schülerleistung gezeichnet werden. Bei beiden Teilfragen wurde der Zusammenhang zwischen der besuchten Schulform und den kognitiven Fähigkeiten mathematischer Schülerleistung detailliert untersucht. Damit wurde ein Schritt dahingehend gemacht, mathematische Schülerleistung aus der Perspektive eines psychologischen Konstrukts besser zu verstehen. Dies ist auch das anvisierte Ziel der Validitätsanalysen im nächsten Kapitel.

## 9 Studie 3: Validität

In diesem Kapitel werden die Beziehungen der kognitiven Fähigkeiten mathematischer Schülerleistung zu motivationalen und soziodemografischen Schülermerkmalen sowie zu Schulnoten untersucht. Diese Schülermerkmale wurden gewählt, um eine möglichst große Bandbreite an häufig untersuchten Außenkriterien abzudecken. Beispielsweise werden die folgenden Fragen behandelt: Sind Jungen besser in Mathematik als Mädchen? Wenn ja, in welchen kognitiven Fähigkeiten der mathematischen Schülerleistung sind diese Unterschiede am deutlichsten? Sind Unterschiede des familiären Hintergrunds mit der mathematikspezifischen Fähigkeit assoziiert? Wie sehr sind Schüler mit hoher mathematikspezifischer Fähigkeit an Mathematik oder am Lesen interessiert, und wie schätzen diese Schüler ihre mathematische oder verbale Begabung ein? Antworten auf diese Fragen betreffen die (externe) Validität der kognitiven Fähigkeiten mathematischer Schülerleistung.

Die Beziehung von Maßen mathematischer Schülerleistung zu den oben aufgeführten Außenkriterien ist bereits in einer Vielzahl von Forschungsarbeiten untersucht worden. Warum sind die Analysen in der vorliegenden Arbeit dennoch wichtig? Wie in Kapitel 6 festgestellt wurde, wurde bei bisherigen Validitätsanalysen der mathematischen Schülerleistung in nahezu allen Studien ein Strukturmodell verwendet, dass dem Standardmodell (vgl. Abb. 16, Modell 5a) entspricht. Diese Studien beziehen sich also vornehmlich auf die generelle mathematische Fähigkeit. Das heißt, Fragen zur externen Validität von kognitiven Fähigkeiten mathematischer Schülerleistung, die in Form eines Nested-Faktormodells spezifiziert werden, sind noch weitestgehend offen (siehe Gustafsson & Snow, 1997, für einige wenige Ausnahmen).

Das hat wichtige Implikationen für die Bewertung und Interpretation der in der Literatur berichteten Validitätskoeffizienten. Folgt man den Annahmen des Nested-Faktormodells (vgl. Abb. 16, Modell 7a), dann gehen in die Varianz der generellen mathematischen Fähigkeit Varianzanteile der allgemeinen kognitiven Fähigkeit und der mathematikspezifischen Fähigkeit „gemischt“ ein. Das ist problematisch, da aufgrund theoretischer Überlegungen und empirischer Befunde erwartet werden kann, dass die allgemeine kognitive Fähigkeit und die mathematikspezifische Fähigkeit unterschiedliche korrelative Beziehungen zu einigen Außenkriterien aufweisen. Dieser Sachverhalt wird in der vorliegenden Arbeit als *differenzielle Validität* bezeichnet. Wenn differenzielle Validitäten vorliegen, bedeutet das gleichzeitig, dass die in der Literatur berichteten Validitätskoeffizienten ein unspezifiziertes Gemisch der Korrelationen zwischen allgemeiner kognitiver Fähigkeit und mathematikspezifischer Fähigkeit darstellen.

Im Rahmen der dritten Forschungsfrage soll daher einerseits versucht werden, bei Verwendung des Standardmodells Befunde aus bisherigen Studien zur externen Validität mathematischer Schülerleistung zu replizieren. Andererseits soll untersucht werden, ob im Vergleich hierzu mit der Konzeptualisierung der kognitiven Fähigkeiten mathematischer Schülerleistung in Form des Nested-Faktormodells ein genaueres Bild der Zusammenhänge zwischen den kognitiven Fähigkeiten mathematischer Schülerleistung und verschiedenen Außenkriterien gezeichnet werden kann.

Zentral für die Hypothesen zur dritten Forschungsfrage ist der Vergleich korrelativer oder regressionsanalytischer Validitätskoeffizienten der generellen mathematischen Fähigkeit ( $M$ ) mit den korrespondierenden Koeffizienten der mathematikspezifischen Fähigkeit ( $M'$ ). Da – entsprechend den Annahmen des Nested-Faktormodells – in die generelle mathematische Fähigkeit die allgemeine kognitive Fähigkeit und die mathematikspezifische Fähigkeit gemischt eingehen,

werden nachfolgend auch die korrelativen Beziehungen zwischen der allgemeinen kognitiven Fähigkeit (bzw. der hierzu äquivalenten fluiden Fähigkeit, Gustafsson, 1984) und soziodemografischen und motivationalen Schülermerkmalen sowie den Schulnoten beleuchtet.<sup>28</sup>

Im Rahmen von Kapitel 9 werden getrennt für die jeweiligen Schülermerkmale zunächst die empirische Befundlage und theoretische Überlegungen dargestellt. Auf dieser Grundlage werden Hypothesen abgeleitet, auf welche Weise die kognitiven Fähigkeiten mathematischer Schülerleistung mit den jeweiligen Außenkriterien assoziiert sind. Daran schließen Abschnitte an, die die verwendete Methodik zur Validitätsanalyse darstellen, die Ergebnisse dokumentieren sowie diskutieren. Abschließend folgt mit Abschnitt 9.6 eine Gesamtdiskussion der dritten Forschungsfrage.

Aus konzeptioneller Sicht ist an dieser Stelle noch anzumerken, dass auch die residualisierte mathematische Fähigkeit  $M'_{HO}$  und die allgemeine kognitive Fähigkeit  $g_{HO}$ , wie sie in einem *Higher-Order*-Modell (Abb. 16, Modell 6a) spezifiziert werden, interessant für die Analysen zur externen Validität sein können. Dies wird angesichts der Überlegung deutlich, dass die mathematikspezifische Fähigkeit im Nested-Faktormodell unabhängig von der allgemeinen kognitiven Fähigkeit und somit in gewissem Sinne ebenfalls „residualisiert“ ist. Gegen eine Verwendung des *Higher-Order*-Modells sprachen jedoch zwei Gründe:

- Der Modell-Fit des Nested-Faktormodells war im Vergleich zum *Higher-Order*-Modell besser. Dies stützte die Favorisierung des Nested-Faktormodells (vgl. Abschnitt 7.2.2; siehe auch McDonald, 1999).
- Es ist zu bedenken, dass der Einfluss (in Form der Faktorladungen) von  $M'_{HO}$  und  $g_{HO}$  auf Maße mathematischer Schülerleistung einer so genannten „Proportionalitätsrestriktion“ unterliegt (Mulaik & Quartetti, 1997; Schmiedek & Li, 2004). Der Einfluss von  $M'_{HO}$  und  $g_{HO}$  auf die manifesten Variablen kann so nicht unabhängig voneinander variieren, da dieser über dieselben Faktorladungen (das sind die Ladungen der Maße mathematischer Schülerleistung auf  $M$ , siehe Abb. 16) vermittelt wird. Dies ist dann von Nachteil, wenn man – wie in dieser Arbeit – die Beziehungen der allgemeinen kognitiven Fähigkeit unabhängig von der mathematikspezifischen Fähigkeit analysieren will.

Aus beiden Gründen wurde in der vorliegenden Arbeit dem Nested-Faktormodell zur Konzeptualisierung der kognitiven Fähigkeiten mathematischer Schülerleistung der Vorzug gegeben. Zur Vervollständigung der Ergebnisse sind im Anhang, aber auch die Korrelationen der untersuchten Schülervariablen mit  $M'_{HO}$  und  $g_{HO}$  dokumentiert. In Tabelle 50 ist auch eine vollständige Übersicht der Korrelationen der Schülermerkmale mit den verbalen Fähigkeiten ( $V$ ,  $V'$  und  $V'_{HO}$ ) enthalten.

---

<sup>28</sup> In der Forschungsliteratur werden in der Regel keine Zusammenhänge zwischen stoffgebietspezifischen Fähigkeiten und Außenkriterien berichtet. Diesem Ansatz folgend werden auch in dieser Arbeit die Zusammenhänge zwischen diesen Schülermerkmalen und mathematischen Teilfähigkeiten *nicht* untersucht.

## 9.1 Geschlechterunterschiede

### 9.1.1 Empirische Befundlage und Hypothesen

Geschlecht ist eines der zentralen soziodemografischen Schülermerkmale, das mit Unterschieden der mathematischen Schülerleistung in Verbindung gebracht wird. Geschlechterunterschiede in mathematischen Fähigkeiten, aber auch in der allgemeinen kognitiven Fähigkeit sind in zahlreichen Einzelstudien dokumentiert. Folgendes Fazit kann mit Blick auf diese kognitiven Fähigkeiten gezogen werden:

*Mathematische Fähigkeiten.* Nach Geary (1996) gibt es keine Geschlechterunterschiede bei den primären mathematischen Fähigkeiten.

Betrachtet man mathematische Schülerleistungen im engeren Sinne (sekundäre mathematische Fähigkeiten sensu Geary), so fanden Hyde, Fennema und Lamon (1990) in ihrer Meta-Analyse bei Studien, die die Gesamtpopulation untersuchten, keinen empirischen Hinweis auf den Vorteil von Jungen, sondern sogar einen leichten Vorsprung der Mädchen (von 0,08 Standardabweichungen). Jedoch kehrte sich das Bild zu Gunsten der Jungen mit zunehmendem Alter um: So erzielten bei 15- bis 18-jährigen Jugendlichen Jungen im Durchschnitt 0,3 Standardabweichungen bessere Leistungen als Mädchen, wobei die Unterschiede am deutlichsten bei mathematischen Problemlöseaufgaben waren (siehe hierzu auch Geary, 1996).

Diese Befundlage wird auch durch die Ergebnisse aus den PISA-Studien aus den Jahren 2000 und 2003 gestützt. In den 32 Staaten, die an der PISA-2000-Studie teilnahmen, hatten bei den 15-jährigen Jugendlichen Jungen im Durchschnitt eine um 0,11 Standardabweichungen (dies entspricht 11 Punkten auf der internationalen PISA-Metrik mit MW 500 und SD 100) bessere Leistung als Mädchen. In Deutschland lag der Vorsprung der 15-jährigen Jungen mit 0,15 Standardabweichungen in ähnlicher Höhe. Bei der Interpretation dieses Befunds ist zu beachten, dass Mädchen in Deutschland tendenziell höhere Bildungsgänge besuchen als Jungen. Betrachtet man die Geschlechterunterschiede innerhalb von Bildungsgängen, so resultierte zum Beispiel bei gymnasialen Bildungsgängen ein Vorteil für die Jungen, der im Durchschnitt bei 0,2 Standardabweichungen lag (Stanat & Kunter, 2001). Ein nahezu identisches Befundmuster zeigte sich auch in der PISA-2003-Studie (Zimmer, Burba & Rost, 2004) und in den Studien BIJU und TIMSS (Hosenfeld, Köller & Baumert, 1999).

*Fluide Fähigkeit und allgemeine kognitive Fähigkeit.* Im Gegensatz zu mathematischen Fähigkeiten belegen einige Studien, dass *keine* Geschlechterunterschiede hinsichtlich der allgemeinen kognitiven Fähigkeit bestehen. Zu diesem Schluss kommen zum Beispiel Jensen (1998) in seinem Standardwerk „The g factor. The science of mental ability“ auf Grundlage einer Reanalyse mehrerer Datensätze sowie die Mitglieder der APA Taskforce, die den Bericht „Intelligence: Knowns and Unknowns“ (Neisser u.a., 1996, S. 91) verfassten. Auch in den Ergebnissen der deutschen Normierungsstichproben des kognitiven Fähigkeitstests werden altersstufenübergreifend keine Geschlechterunterschiede dokumentiert (Heller & Perleth, 2000). Darüber hinaus ist festzuhalten, dass Tests zur Messung der allgemeinen kognitiven Fähigkeit auch häufig so konstruiert werden, dass keine Geschlechterdifferenzen entstehen (Neisser u.a., 1996, S. 91).

Auf Grundlage der dokumentierten Befundlage können die folgenden Hypothesen zu Geschlechterunterschieden bei den kognitiven Fähigkeiten mathematischer Schülerleistung abgeleitet werden. Hypothesen werden in dieser Arbeit in Form korrelativer Beziehungen formuliert. Im Fall



von Geschlechterunterschieden werden also Mittelwertunterschiede als punkt-biseriale Korrelationen parametrisiert. Damit die Hypothesen der dritten Forschungsfrage klar von denen der zweiten Forschungsfrage unterschieden werden können, werden die Hypothesen fortlaufend nummeriert.

*Hypothese 3a.* In Anlehnung an Jensen wird davon ausgegangen, dass keine Geschlechterunterschiede hinsichtlich der allgemeinen kognitiven Fähigkeit ( $g$ ) bestehen ( $r_{g, \text{Geschlecht}} = 0$ ).

*Hypothese 3b.* Aufgrund der postulierten Äquivalenz von allgemeiner kognitiver Fähigkeit und fluider Fähigkeit ( $Gf$ ) sollten ebenfalls keine Geschlechterunterschiede hinsichtlich der fluiden Fähigkeit bestehen ( $r_{Gf, \text{Geschlecht}} = 0$ ).

*Hypothese 3c.* Entsprechend der Befunde der Meta-Analyse von Hyde und Kollegen wie auch der empirischen Ergebnisse im Rahmen von *Large-Scale*-Studien kann davon ausgegangen werden, dass Jungen durchschnittlich bessere Leistungen bei der generellen mathematischen Fähigkeit ( $M$ ) erzielen als Mädchen. In dieser Arbeit werden Jungen mit 1 und Mädchen mit 0 codiert. Daher ist zu erwarten, dass gilt:  $r_{M, \text{Geschlecht}} > 0$ .

*Hypothese 3d.* Geschlechterunterschiede sind bei mathematischen Problemlöseaufgaben im Mittel zu Gunsten der Jungen ausgeprägt (Geary, 1996; Hyde u.a., 1990). Da die mathematikspezifische Fähigkeit in dieser Arbeit als Fähigkeit zum mathematischen Problemlösen interpretiert wird, sollten Jungen im Durchschnitt bessere Leistungen bei der mathematikspezifischen Fähigkeit ( $M'$ ) erreichen als Mädchen ( $r_{M', \text{Geschlecht}} > 0$ ).

*Hypothese 3e.* Geschlechterunterschiede zu Gunsten der Jungen sollten bei der mathematikspezifischen Fähigkeit größer sein als bei der generellen mathematischen Fähigkeit ( $r_{M, \text{Geschlecht}} < r_{M', \text{Geschlecht}}$ ). Hierfür sprechen drei Gründe:

- In die generelle mathematische Fähigkeit gehen Varianzanteile der mathematikspezifischen Fähigkeit und der allgemeinen kognitiven Fähigkeit gemischt ein. Da keine Geschlechterunterschiede für  $Gf$  bzw.  $g$  erwartet werden, führt dies dazu, dass die Geschlechterunterschiede der generellen mathematischen Fähigkeit geringer sind als bei der mathematikspezifischen Fähigkeit.
- Bei Kontrolle der Bildungsgänge oder Schulformen nehmen Geschlechterunterschiede zu Gunsten der Jungen zu. Da die allgemeine kognitive Fähigkeit ein zentraler Prädiktor für den besuchten Bildungsgang oder die besuchte Schulform ist, werden bei Konstanthalten der allgemeinen kognitiven Fähigkeit folglich teilweise bildungsgang- oder schulformspezifische Unterschiede konstant gehalten.
- Geschlechterunterschiede sind bei mathematischen Problemlöseaufgaben durchschnittlich stärker ausgeprägt (Hyde u.a., 1990). Da die mathematikspezifische Fähigkeit in dieser Arbeit als Fähigkeit zum mathematischen Problemlösen interpretiert wird, sollten Jungen im Durchschnitt bessere Leistungen bei der mathematikspezifischen Fähigkeit erreichen als bei der generellen mathematischen Fähigkeit.

### 9.1.2 Methode

*Stichprobe.* Alle Analysen im Rahmen der dritten Forschungsfrage wurden anhand der Daten der Analysestichprobe mit 29.386 Schülern durchgeführt. Eine Beschreibung dieser Stichprobe befindet sich in Abschnitt 7.1.2 (Tab. 3).

*Manifeste Variablen.* Indikatoren der kognitiven Fähigkeiten waren bei allen Analysen zur externen Validität die WLE-Scores zur Messung der mathematischen Fähigkeit, der verbalen Fähigkeit und der fluiden Fähigkeit, die auch bei den Analysen im Rahmen der ersten und zweiten Forschungsfrage verwendet wurden (für eine nähere Beschreibung siehe Abschnitt 7.2.2).

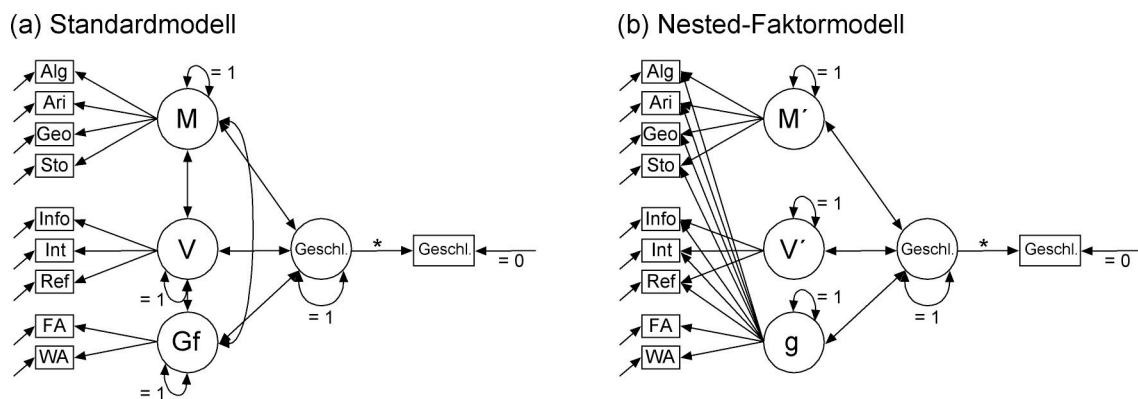
Das Geschlecht wurde anhand einer so genannten Trackingvariable aus dem PISA-2000-Datensatz bestimmt (Kunter u.a., 2002, S. 24). Für die Analysen in der vorliegenden Arbeit wurden Jungen mit 1 und Mädchen mit 0 codiert. Eine positive Korrelation indizierte somit einen Leistungsvorteil zu Gunsten der Jungen. Für 60 Schüler lag keine Angabe zum Geschlecht vor.

*Analyseverfahren.* Die Analysen aller korrelativen Beziehungen basieren auf dem Standardmodell (Modell 5a, siehe Abb. 16 und Abschnitt 7.2.2) und dem Nested-Faktormodell (Modell 7a, siehe Abb. 16 und Abschnitt 7.2.2). Damit die Hypothesen zu korrelativen Zusammenhängen bestmöglich analysiert werden konnten, wurde folgendermaßen vorgegangen:

- (a) Um simultan das Problem fehlender Werte und die hierarchische Datenstruktur zu berücksichtigen, wurden alle Modellparameter mit dem Algorithmus MLR und dem Modul *complex* des Programms Mplus 3.01 (Muthén & Muthén, 1998–2004b) geschätzt.

Wenn manifeste Variablen, wie zum Beispiel das Geschlecht, als exogene Variablen im Modell spezifiziert werden und damit weder von anderen manifesten noch latenten Variablen abhängig sind, kann Mplus das Problem fehlender Werte programmbedingt zunächst nicht lösen. Um dies zu umgehen, wurde ein latenter Geschlechterfaktor in das Modell aufgenommen. Die manifeste Geschlechtervariable war damit vom latenten „Geschlechterfaktor“ abhängig und wurde so zur endogenen Variablen. Zur Spezifikation der latenten Geschlechtervariable wurde die latente Varianz auf 1,0 und der Residualterm der manifesten Geschlechtervariable auf Null fixiert. Die frei geschätzte Faktorladung repräsentierte damit die Standardabweichung der manifesten Geschlechtervariablen.

- (b) Die Hypothesen 3a bis 3d beziehen sich auf die Korrelation zwischen kognitiven Fähigkeiten und einer Geschlechtervariablen. Um die Hypothesen zu testen, müssen sich die spezifizierten Modellparameter nicht wie ansonsten üblich auf die Kovarianzen, sondern auf die Korrelationen beziehen. Hierzu wurden zusätzlich zur latenten Geschlechtervariablen die latenten Variablen im Standardmodell (Abb. 26a) wie auch im Nested-Faktormodell (Abb. 26b) auf 1,0 fixiert, um die Modelle zu identifizieren. Durch dieses Vorgehen wird eine Kovarianz zwischen zwei Variablen geschätzt, deren Varianz jeweils 1,0 ist. Dies entspricht exakt der Korrelation zwischen diesen beiden Variablen.



M = generelle mathematische Fähigkeit, V = generelle verbale Fähigkeit, Gf = fluide Fähigkeit, M' = mathematikspezifische Fähigkeit, V' = spezifische verbale Fähigkeit, g = allgemeine kognitive Fähigkeit, Geschl. = Mädchen wurden mit 0, Jungen wurden mit 1 codiert. Der Stern kennzeichnet den frei geschätzten Parameter, der die Standardabweichung der Geschlechtervariablen repräsentierte.

Abbildung 26: Modellspezifikation zur Prüfung der Geschlechterunterschiede

Zur inferenzstatistischen Prüfung wurden  $\chi^2$ -Differenzentests verwendet. Hierzu wurden die  $\chi^2$ -Goodness-of-Fit-Statistiken eines Modells, in dem die Korrelation mit der latenten Geschlechtervariable frei geschätzt wurde, verglichen mit der  $\chi^2$ -Goodness-of-Fit-Statistik eines Modells, in dem diese Korrelation auf Null (Hypothesen 3a, 3b, 3c und 3d) fixiert wurde. Zur Prüfung von Hypothese 3e zur Ungleichheit der Geschlechterunterschiede in M und M' wurden die Korrelationen im Nested-Faktormodell auf die korrespondierende Korrelation im Standardmodell fixiert. Eine signifikante Verschlechterung des  $\chi^2$ -Goodness-of-Fit-Werts für das Nested-Faktormodell indizierte die Ungleichheit der Korrelationen.

Bei allen inferenzstatistischen Tests wurde ein Signifikanzniveau von  $p < .05$  gesetzt.

- (c) Es besteht die Gefahr, dass sich die Modellparameter der latenten Variablen zur Messung der kognitiven Fähigkeiten im Standardmodell und im Nested-Faktormodell bei Analyse verschiedener Außenkriterien verändern. Somit wären die Modelle nicht invariant, wenn unterschiedliche Schülermerkmale analysiert werden. Um dies zu verhindern, wurden in einem ersten Schritt die Modellparameter des Standardmodells und des Nested-Faktormodells (mit jeweils auf 1,0 fixierten latenten Varianzen) geschätzt. Die Geschlechtervariable war dabei *nicht* als manifeste Variable im Modell enthalten. Weiterhin wurden zur Identifikation des Faktors für die fluide Fähigkeit die Faktorladungen der Figurenanalogien und der Wortanalogien auf den gleichen Wert restringiert.

In einem zweiten Schritt wurden zur Analyse der Geschlechterunterschiede alle Modellparameter (Faktorladungen, Residualterme, Intercepts, latente Varianzen und latente Korrelationen) der beiden Strukturmodelle auf die Werte fixiert, die im ersten Schritt ermittelt wurden. Die Korrelationen zwischen den latenten Variablen und der latenten Geschlechtervariablen wie auch die Faktorladung der manifesten Geschlechtervariablen wurden frei geschätzt.

*Kontrolle der Schulformzugehörigkeit.* Aufgrund der disproportionalen Verteilung der Jungen und Mädchen auf die unterschiedlichen Schulformen (vgl. Tab. 3) wurden die Geschlechterunterschiede auch bei Kontrolle der Schulformzugehörigkeit untersucht.

Prinzipiell würde es sich anbieten, eine Regression aller latenten Variablen im Standardmodell oder Nested-Faktormodell (inkl. der latenten Geschlechtervariable) auf Dummyvariablen zu berechnen, die die Schulformzugehörigkeit repräsentieren. Dieser Weg wurde jedoch nicht eingeschlagen: Bei einer Regression korrelieren die latenten Variablen in Abhängigkeit von den Regressionsgewichten der Prädiktoren miteinander. Dies impliziert zum Beispiel, dass im Nested-Faktormodell die Faktoren für die kognitiven Fähigkeiten nicht mehr orthogonal zueinander stehen (siehe hierzu Schmiedek & Li, 2004).

Aus diesem Grund wurde eine andere Analysestrategie gewählt, die in drei Schritten vollzogen wurde. Im ersten Schritt wurden fehlende Werte für alle Schülermerkmale, die im Rahmen der dritten Forschungsfrage untersucht wurden, und Leistungsindikatoren auf Grundlage von EM-Parametern mit dem Programm Norm (Schafer, 2000) imputiert. Somit lag für jeden Schüler der beste Punktschätzer für die fehlenden Daten vor (Graham, Cumsille & Elek-Fisk, 2003). Im Imputationsmodell wurde die Schulformzugehörigkeit als Dummyvariable berücksichtigt. Weiterhin wurden fehlende Werte bei der Geschlechtervariable auf die nächste ganze Zahl (hier 0 oder 1) gerundet.

Im zweiten Schritt wurde die Schulformzugehörigkeit bei diesen EM-imputierten Daten kontrolliert. Hierzu wurde eine lineare Regression berechnet: Abhängige Variablen waren die neun WLE-Scores zur Messung der kognitiven Fähigkeiten sowie die weiteren Variablen zur Messung der motivationalen und soziodemografischen Schülermerkmale und Schulnoten. Unabhängige Variablen waren vier Dummyvariablen, die die Zugehörigkeit zu einer der fünf Schulformen (vgl. Tab. 3) indizierten. Die Hauptschule war dabei die Referenzschulform, das heißt, Hauptschüler hatten bei allen vier Dummyvariablen den Wert Null. Die unstandardisierten Regressionsresiduen repräsentierten dann die Werte eines Schülers, die unabhängig von seiner Schulformzugehörigkeit waren.

Im dritten Schritt wurden dann die unstandardisierten Regressionsresiduen als manifeste Variablen verwendet. Die weiteren Analysen wurden analog zur oben beschriebenen Vorgehensweise durchgeführt. Die Kontrolle der Schulformzugehörigkeit war nicht zentral zur Prüfung der postulierten Hypothesen, die sich auf die Gesamtpopulation konzentrierten. Daher werden die Ergebnisse unter Kontrolle der Schulformzugehörigkeit nur deskriptiv berichtet und nicht inferenzstatistisch getestet. Hierfür sprach auch, dass bei Verwendung von EM-imputierten Daten die Standardfehler nicht korrekt sind (z.B. Enders, 2001; Graham u.a., 2003).

### 9.1.3 Ergebnisse

Die Ergebnisse zu Geschlechterunterschieden werden analog zur Reihenfolge der Hypothesen berichtet und sind in Tabelle 26 dokumentiert. Eine grafische Übersicht zu den zentralen Ergebnissen der dritten Forschungsfrage bieten die Abbildungen 30a und 30b in Abschnitt 9.6. In Anlehnung an die Bewertung von Effektstärken, die Cohen (1992) vorschlägt, werden Korrelationen in der Höhe von .10 als „klein“ oder „gering“, Korrelationen in der Höhe von .30 als „mittel“ und Korrelationen größer .50 als „groß“ bezeichnet.

*Geschlechterunterschiede in g und Gf.* Jungen und Mädchen unterschieden sich nur geringfügig in der allgemeinen kognitiven Fähigkeit wie auch hinsichtlich der fluiden Fähigkeit. Jedoch war der leichte Leistungsvorsprung der Mädchen in der allgemeinen kognitiven Fähigkeit wie auch in der fluiden Fähigkeit statistisch signifikant von Null verschieden.

Die Hypothesen 3a und 3b wurden somit aus inferenzstatistischer Sicht nicht von den Daten gestützt. Dabei ist aber zu bedenken, dass üblicherweise in der psychologischen Forschung Korrelationen in der Höhe von  $-.05$  eher im Sinne eines Nulleffekts als im Sinne eines psychologisch bedeutsamen Effekts interpretiert werden.

Der leichte Leistungsvorsprung der Mädchen konnte zudem durch die disproportionale Verteilung auf die Schulformen erklärt werden. Hielt man die Schulformzugehörigkeit in den manifesten Variablen konstant, verschwand die leicht negative Korrelation (Tab. 26).

*Geschlechterunterschiede in M und M'*. Jungen erzielten geringfügig bessere Leistungen in der generellen mathematischen Fähigkeit als Mädchen (Hypothese 3c). Hingegen kann der Leistungsvorsprung in der mathematikspezifischen Fähigkeit als mittlerer bis starker „Effekt“ interpretiert werden (Hypothese 3d). Die Geschlechterunterschiede waren also bei beiden mathematischen Fähigkeiten nicht gleich stark ausgeprägt (Hypothese 3e): Jungen waren wesentlich besser bei der mathematikspezifischen Fähigkeit als bei der generellen mathematischen Fähigkeit.

Alle Hypothesen zu den mathematischen Fähigkeiten wurden auch (zusätzlich zu den deskriptiven Befunden) inferenzstatistisch gestützt.

Kontrollierte man nun zusätzlich für die Schulformzugehörigkeit, nahm der Leistungsvorsprung zu Gunsten der Jungen bei der generellen mathematischen Fähigkeit deutlich zu. Die Geschlechterdifferenz in der mathematikspezifischen Fähigkeit blieb unverändert und war im Vergleich zur generellen mathematischen Fähigkeit immer noch stärker ausgeprägt.

Tabelle 26: Korrelationen ( $r$ ) zwischen kognitiven Fähigkeiten (KF) und Geschlecht sowie korrespondierende  $\chi^2$ -Differenzentests ( $\Delta\chi^2$ ) zur Prüfung der Hypothesen (Hpt) zu Geschlechterunterschieden

Hpt	KF	$r$	$r_{\text{Schulform}}$	Test	$\chi^2$ (df = 61)	SF	$\Delta\chi^2$
3a	g	$-.05$	$.01$	$r_{g, \text{Geschlecht}} = 0$	700	1,556	40,0*
3b	Gf	$-.04$	$.00$	$r_{Gf, \text{Geschlecht}} = 0$	1.256	1,556	36,0*
3c	M	$.13$	$.24$	$r_{M, \text{Geschlecht}} = 0$	1.488	1,558	308,5*
3d	M'	$.42$	$.42$	$r_{M', \text{Geschlecht}} = 0$	1.531	1,556	1121,1*
3e	M & M'			$r_{M, \text{Geschlecht}} = r_{M', \text{Geschlecht}}$	1.093	1,556	551,9*

\*  $p < .05$ .

$r_{\text{Schulform}}$  = Korrelation bei Kontrolle der Schulformzugehörigkeit, SF = Skalierungsfaktor. Zur Berechnung der  $\chi^2$ -Differenz ( $\Delta\chi^2$ ) wurde die Korrekturformel aus dem technischen Anhang von Mplus verwendet (Muthén & Muthén, 1998–2004a, S. 22). Alle  $\chi^2$ -Differenzen wurden mit df = 1 getestet. Der  $\chi^2$ -Goodness-of-Fit-Testwert des unrestringierten Standardmodells, mit dem die jeweilige Parameterrestriktion verglichen wurde, betrug  $\chi^2 = 1.224$  mit df = 60 und SF = 1.562 (Nested-Faktormodell:  $\chi^2 = 667$ , df = 60, SF = 1.562). Fähigkeiten im Standardmodell: Gf = fluide Fähigkeit, M = generelle mathematische Fähigkeit. Fähigkeiten im Nested-Faktormodell: g = allgemeine kognitive Fähigkeit, M' = mathematikspezifische Fähigkeit.

#### 9.1.4 Diskussion

Die wichtigsten Ergebnisse zu den Geschlechterunterschieden lassen sich wie folgt zusammenfassen:

- Jungen und Mädchen unterschieden sich kaum in der allgemeinen kognitiven Fähigkeit und fluiden Fähigkeit (Hypothese 3a und 3b). Die relativ geringen Unterschiede verschwanden bei Kontrolle der Schulformzugehörigkeit. Weiterhin waren Geschlechterunterschiede in der allgemeinen kognitiven Fähigkeit und fluiden Fähigkeit gleich groß.
- Mit Blick auf die mathematischen Fähigkeiten ist festzuhalten, dass Jungen sowohl bei der generellen mathematischen Fähigkeit (Hypothese 3c) wie auch bei der mathematikspezifischen Fähigkeit (Hypothese 3d) bessere Leistungen erzielten. Jedoch waren die Geschlechterunterschiede wesentlich stärker bei der mathematikspezifischen Fähigkeit zu Gunsten der Jungen ausgeprägt (Hypothese 3e). Bei Kontrolle der Schulformzugehörigkeit vergrößerten sich die Unterschiede der generellen mathematischen Fähigkeit zu Gunsten der Jungen, wohingegen sie bei der mathematikspezifischen Fähigkeit konstant blieben.

Diese Ergebnisse stützen die Annahme, dass bei Verwendung des Standardmodells in die Geschlechterunterschiede der generellen mathematischen Fähigkeit Unterschiede in der allgemeinen kognitiven Fähigkeit und Unterschiede in der mathematikspezifischen Fähigkeit gemischt eingehen. Mathematikspezifische Geschlechterunterschiede werden damit unterschätzt (siehe hierzu auch Wittmann, 2005).

Kontrollierte man zusätzlich für die Schulformzugehörigkeit, nahm wie erwartet (Hosenfeld u.a., 1999; Stanat & Kunter, 2001; Zimmer u.a., 2004) der Leistungsvorsprung zu Gunsten der Jungen bei der generellen mathematischen Fähigkeit zu. Der Grund hierfür ist, dass durch Kontrolle der Schulformzugehörigkeit Unterschiede in der allgemeinen kognitiven Fähigkeit zwischen den Schulen reduziert wurden. Somit nahm der proportionale Anteil der mathematikspezifischen Fähigkeit an der Gesamtvarianz der generellen mathematischen Fähigkeit zu. Damit vergrößerten sich die Geschlechterunterschiede in der generellen mathematischen Fähigkeit.

Hingegen blieben die Geschlechterunterschiede in der mathematikspezifischen Fähigkeit auch nach Kontrolle der Schulformzugehörigkeit nahezu konstant: Unabhängig von der besuchten Schulform weisen Jungen also bei gleicher allgemeiner kognitiver Fähigkeit eine deutlich stärker ausgeprägte mathematikspezifische Fähigkeit als Mädchen auf. Die deutlichen Geschlechterunterschiede bei der mathematikspezifischen Fähigkeit zeigen damit auf, wo die größten Schwierigkeiten der Mädchen im Fach Mathematik liegen könnten. Folgt man der Interpretation der mathematikspezifischen Fähigkeit (vgl. Abschnitt 7.3), so bestehen die Geschlechterunterschiede in der spezifischen Fähigkeit zum mathematischen Problemlösen. Diese Interpretation wird auch durch die Meta-Analyse von Hyde und Kollegen (1990) gestützt (siehe auch Geary, 1996): Bei Highschool-Schülern waren Geschlechterunterschiede zu Gunsten der Jungen am stärksten bei mathematischen Problemlöseaufgaben ausgeprägt. Instruktionale Maßnahmen zur Vermittlung von mathematischen Problemlösetechniken (vgl. Abschnitt 5.3; zusammenfassend z.B. De Corte, Greer & Verschaffel, 1996) könnten also möglicherweise insbesondere die Leistung der Mädchen bei mathematischen Problemlöseaufgaben deutlich verbessern.

## 9.2 Familiärer Hintergrund

### 9.2.1 Empirische Befundlage und Hypothesen

Neben Geschlechterunterschieden in der mathematischen Schülerleistung wird als weiteres soziodemografisches Schülermerkmal häufig der Zusammenhang zwischen der mathematischen Schülerleistung und dem familiären Hintergrund untersucht. Aus psychologischer Perspektive können Indikatoren des familiären Hintergrunds als (distales) Maß der Fördermöglichkeiten von Kindern und Jugendlichen in den Familien betrachtet werden.

Förderungsmöglichkeiten von Kindern und Jugendlichen in der Familie sind vielfältig. Familien unterscheiden sich dabei nach Helmke und Weinert (1997) in der Stimulation kognitiver Aktivitäten (z.B. Vorlesen oder ob Eltern ihren Kindern kognitiv anregendes Spielzeug geben), im instruktionalen Verhalten (z.B. inwiefern Eltern ihre Kinder an Frühprogrammen in Kindergärten teilnehmen lassen oder in der Art und Weise, wie sie das Bearbeiten der Hausaufgaben unterstützen), in der Förderung schulischer Lernmotivation (z.B. durch Bildungsaspirationen, Leistungsdruck, Belohnung und Bestrafung) und inwiefern ihr eigenes intellektuelles Verhalten den Kindern Vorbild für kognitive Aktivitäten ist (z.B. Lernen am Modell aufgrund der Beobachtung, dass Mutter oder Vater gerne lesen). Ein nicht zu vernachlässigender Faktor ist auch das gesundheitsrelevante Verhalten der Eltern, die zum Beispiel durch Ernährung und Hygiene notwendige Bedingungen für eine positive kognitive Entwicklung schaffen (Bradley & Corwyn, 2002).

In der Lehr-Lern-Forschung ist es fast nie möglich, alle diese psychologisch wirksamen Faktoren detailliert zu erfassen. Daher bedient man sich proximaler Variablen, die mit dem lernförderlichen Elternverhalten zusammenhängen (zusammenfassend z.B. Baumert & Schümer, 2001; Bradley & Corwyn, 2002). Mit Baumert und Schümer kann man drei interdependente Aspekte des familiären Hintergrunds unterscheiden:

- (a) Die *sozioökonomische Stellung* einer Familie ist mit einer ganzen Reihe von lernförderlichen Mechanismen und Zugang zu lernförderlichen, finanziell erschließbaren Ressourcen verbunden (siehe auch Bradley & Corwyn, 2002). So wurde bei PISA 2000 die sozioökonomische Stellung der Eltern über ihre Berufstätigkeit (als ein Indikator für das finanzielle Einkommen der Eltern) und den Besitz materieller Güter (z.B. Fernseher, Computer, Geschirrspülmaschine) erfasst (Baumert & Schümer, 2001). Während die sozioökonomische Stellung in erster Linie durch ökonomische Ressourcen (finanzielles Kapital) charakterisiert ist, können hiervon zwei weitere Kapitale unterschieden werden.
- (b) *Kulturelles Kapital* sensu Bourdieu (z.B. 1983) umfasst alle Ressourcen, die grundlegend für die Teilhabe an einer bürgerlichen Gesellschaft sind. Zum kulturellen Kapital gehören einerseits strukturelle Aspekte wie Kunstwerke oder Bücher sowie Bildungszertifikate und akademische Titel. Andererseits gehören hierzu funktionale Aspekte wie Wertorientierungen, Einstellungen und Kompetenzen (Baumert & Schümer, 2001).
- (c) *Soziales Kapital* sensu Coleman (1988) zeichnet sich durch Zugang zu sozialen Netzwerken aus, was im Gegenzug den Austausch von Informationen und die Entwicklung von netzwerkspezifischen Normen und Wertorientierungen fördert. Baumert und Schümer unterscheiden wiederum strukturelle Aspekte sozialen Kapitals (z.B. eine intakte Familienstruktur) von funktionalen Aspekten (z.B. Kommunikation innerhalb der Familie).

Meist werden die drei Beschreibungsmerkmale der sozialen Herkunft eines Schülers in empirischen Untersuchungen nicht differenziert erfasst, sondern global unter den Begriff „SES“ subsumiert. Ein wichtiges empirisches Ergebnis ist in diesem Zusammenhang, dass höhere Werte des SES (gemessen durch ein oder mehrere Indikatoren sozioökonomischer Stellung, kulturellen Kapitals oder sozialen Kapitals) mit höheren Leistungen in der generellen mathematischen Fähigkeit ( $r = .20$ ) einhergehen (White, 1982, Tab. 5).

Im Vergleich zur mathematischen Fähigkeit scheint der Zusammenhang zwischen SES und allgemeiner kognitiver Fähigkeit stärker zu sein. Jensen (1998) berichtet, dass die durchschnittliche Korrelation zwischen SES und allgemeiner kognitiver Fähigkeit zwischen .30 und .40 liegt. Einen ähnlichen Range für die Korrelation zwischen SES und allgemeiner kognitiver Fähigkeit ( $.30 \leq r \leq .35$ ) gibt auch Turkheimer (1994, S. 993) an.

Was kann nun für den Zusammenhang der kognitiven Fähigkeiten mathematischer Schülerleistung mit Variablen des familiären Hintergrunds erwartet werden? Bei der Ableitung von Hypothesen werden die drei Kapitale nicht differenziert betrachtet, sondern ebenfalls global unter den Begriff „SES“ subsumiert.

*Hypothese 4a und 4b.* Die referierten Arbeiten von Jensen und Turkheimer lassen erwarten, dass Indikatoren des familiären Hintergrunds (im Sinne des SES) und die allgemeine kognitive Fähigkeit bzw. die fluide Fähigkeit positiv interkorrelieren (Hypothese 4a:  $r_{g,SES} > 0$ ; Hypothese 4b:  $r_{GF,SES} > 0$ ).

*Hypothese 4c.* Wie die Meta-Analyse von White und viele empirische Einzelbefunde (z.B. Baumert & Schümer, 2001, 2002) nahe legen, ist ein positiver Zusammenhang zwischen der generellen mathematischen Fähigkeit und SES zu erwarten ( $r_{M,SES} > 0$ ).

*Hypothese 4d.* Was ist für den Zusammenhang zwischen SES und der mathematikspezifischen Fähigkeit anzunehmen? Prinzipiell gilt, dass das Wissen und die Fähigkeiten, die zur Lösung von Aufgaben zur Messung mathematischer Schülerleistung benötigt werden, primär in der Schule erworben werden (vgl. Geary, 1995; Köller & Baumert, 2002). Folgt man diesem Argument, dann sollte die mathematikspezifische Fähigkeit von lernförderlichen Aspekten des familiären Hintergrunds unabhängig sein. Daher ist davon auszugehen, dass  $M'$  nicht mit Maßen des SES korreliert ( $r_{M',SES} = 0$ ).

*Hypothese 4e.* Folgt man der Argumentation von Hypothese 5c und 5d, ist davon auszugehen, dass die Korrelation von SES mit  $M$  größer ist als die Korrelation von SES mit  $M'$  ( $r_{M,SES} > r_{M',SES}$ ).

## 9.2.2 Methode

*Manifeste Variablen.* Die Beziehung der kognitiven Fähigkeiten zu Unterschieden des familiären Hintergrunds wurde hinsichtlich von zwei Indikatoren untersucht. Der erste Indikator war die *Anzahl der Bücher* (Anz.-B) in den Familien. Auf die Frage „Wie viele Bücher habt Ihr zu Hause?“ konnten die Schüler eine von sechs Antwortkategorien wählen (1 = maximal 10 Bücher, 2 = zwischen 11 und 50 Bücher, 3 = zwischen 51 und 100 Bücher, 4 = 101 bis 250 Bücher, 5 = 251 bis 500 Bücher, 6 = über 500 Bücher). Die Anzahl der Bücher ist ein Maß für den strukturellen Aspekt des kulturellen Kapitals einer Familie (vgl. Abschnitt 5.2.3; siehe Baumert & Schümer, 2001). Von 734 Schülern fehlten die Angaben zur Anzahl der Bücher.



Als zweiter Indikator wurde der höchste Wert des *International Socio-Economic Index of Occupational Status* (ISEI, Ganzeboom, de Graaf, Treiman & de Leeuw, 1992) in einer Familie herangezogen. Der ISEI wurde gebildet, indem die Angaben zum ausgeübten Beruf von Vater und Mutter nach dem *International Standard Classification of Occupations* (ISCO-88, International Labor Office, 1990) vercodet wurden. Als Maß für die sozioökonomische Stellung, die mit dem Ausüben eines Berufs verbunden ist, wurde der ISEI als Standardindikator bei PISA verwendet (Baumert & Schümer, 2001). Für 1.561 Schüler lag kein Wert des ISEI vor.

*Analyseverfahren.* Die statistischen Analysen erfolgten analog zur Untersuchung der Geschlechterunterschiede. In zwei getrennten Analysen wurden die korrelativen Beziehungen der kognitiven Fähigkeiten im Standardmodell und im Nested-Faktormodell zur Anzahl der Bücher sowie dem ISEI untersucht. Hierzu wurden wiederum latente Variablen für die beiden Indikatoren des familiären Hintergrunds analog zur Analyse der Geschlechterunterschiede gebildet (vgl. Abb. 26a und 26b).

*Kontrolle der Schulformzugehörigkeit.* Die korrelativen Beziehungen zwischen Indikatoren des familiären Hintergrunds und den kognitiven Fähigkeiten wurden auch bei Kontrolle der Schulformzugehörigkeit analysiert (siehe Abschnitt 9.1.2 für die Beschreibung der Methodik). Der Grund hierfür ist, dass Schüler mit höheren Werten auf den Indikatoren des familiären Hintergrunds tendenziell höhere Schulformen besuchen (Baumert & Schümer, 2001). Eine zusätzliche Kontrolle für diese disproportionale Verteilung der Schülerschaft ermöglichte, ein klareres Bild über die Zusammenhänge zwischen Indikatoren des familiären Hintergrunds und den kognitiven Fähigkeiten zu zeichnen.

### 9.2.3 Ergebnisse

Die Ergebnisse werden entsprechend der Reihenfolge der Hypothesen dargestellt und sind in Tabelle 27 getrennt für die beiden Indikatoren des familiären Hintergrunds dokumentiert (siehe auch Abb. 30a und 30b in Abschnitt 9.6 für eine grafische Übersicht der Validitätskoeffizienten).

*Zusammenhang des familiären Hintergrunds mit  $g$  und  $Gf$ .* Beide Indikatoren des familiären Hintergrunds korrelierten mittel (in Bezug auf den ISEI) bzw. mittel bis stark (in Bezug auf die Anzahl an Büchern) mit der fluiden Fähigkeit wie auch mit der allgemeinen kognitiven Fähigkeit.

Wie zu erwarten, verschlechterte sich der globale Modell-Fit deutlich, als diese Korrelationen auf Null fixiert wurden (Tab. 27). Jedoch resultierten bei Anwendung der empfohlenen Korrekturformel (Muthén & Muthén, 1998–2004a, S. 22) negative Werte für die  $\chi^2$ -Differenzen, die nicht interpretiert werden konnten (Satorra & Bentler, 2001). Allerdings legten die deskriptiven Werte wie auch die 95%-Konfidenzintervalle der jeweiligen Korrelationen (siehe Abb. 30a) nahe, dass diese Korrelationen auch statistisch signifikant von Null verschieden waren. Die Hypothesen 4a und 4b wurden somit auf jeden Fall deskriptiv (und mit Einschränkungen auch inferenzstatistisch) gestützt.

Tabelle 27: Korrelationen ( $r$ ) zwischen kognitiven Fähigkeiten (KF) und Indikatoren des familiären Hintergrunds sowie korrespondierende  $\chi^2$ -Differenzentests ( $\Delta\chi^2$ ) zur Prüfung der Hypothesen (Hpt)

Hpt	KF	$r$	$r_{\text{Schulform}}$	Test	$\chi^2$ (df = 61)	SF	$\Delta\chi^2$
<i>Anzahl an Büchern</i>							
4a	g	.41	.24	$r_{g, \text{Anz.-B}} = 0$	2.999	1,523	-4.608,9
4b	Gf	.40	.23	$r_{Gf, \text{Anz.-B}} = 0$	3.300	1,526	-5.486,6
4c	M	.38	.20	$r_{M, \text{Anz.-B}} = 0$	3.411	1,522	-4.140,4
4d	M'	.01	.00	$r_{M', \text{Anz.-B}} = 0$	517	1,554	0,8
4e	M & M'			$r_{M, \text{Anz.-B}} = r_{M', \text{Anz.-B}}$	1.272	1,552	123,9*
<i>ISEI</i>							
4a	g	.35	.15	$r_{g, \text{ISEI}} = 0$	2.202	1,525	-4.976,4
4b	Gf	.33	.14	$r_{Gf, \text{ISEI}} = 0$	2.463	1,528	-6.756,5
4c	M	.34	.15	$r_{M, \text{ISEI}} = 0$	3.411	1,522	-5.259,0
4d	M'	.05	.04	$r_{M', \text{ISEI}} = 0$	525	1,550	17,0*
4e	M & M'			$r_{M, \text{ISEI}} = r_{M', \text{ISEI}}$	960	1,548	776,6*

\*  $p < .05$ .

$r_{\text{Schulform}}$  = Korrelation bei Kontrolle der Schulformzugehörigkeit, SF = Skalierungsfaktor. Zur Berechnung der  $\chi^2$ -Differenz ( $\Delta\chi^2$ ) wurde die Korrekturformel aus dem technischen Anhang von Mplus verwendet (Muthén & Muthén, 1998–2004a, S. 22). Alle  $\chi^2$ -Differenzen wurden mit  $df = 1$  getestet. Der  $\chi^2$ -*Goodness-of-Fit*-Testwert des unrestringierten Standardmodells, mit dem die jeweilige Parameterrestriktion hinsichtlich der Anzahl an Büchern (Anz.-B) verglichen wurde, betrug  $\chi^2 = 997$  mit  $df = 60$  und  $SF = 1.562$  (Nested-Faktormodell:  $\chi^2 = 514$ ,  $df = 60$ ,  $SF = 1.562$ ). Der korrespondierende Wert im Standardmodell für die Analysen zum ISEI war  $\chi^2 = 976$  mit  $df = 60$  und  $SF = 1.559$  (Nested-Faktormodell:  $\chi^2 = 510$ ,  $df = 60$ ,  $SF = 1.559$ ). Fähigkeiten im Standardmodell: Gf = fluide Fähigkeit, M = generelle mathematische Fähigkeit. Fähigkeiten im Nested-Faktormodell: g = allgemeine kognitive Fähigkeit, M' = mathematikspezifische Fähigkeit.

Kontrollierte man für die Schulformzugehörigkeit, verringerten sich bei beiden Indikatoren die Zusammenhänge mit der fluiden Fähigkeit bzw. mit der allgemeinen kognitiven Fähigkeit deutlich. Die korrelativen Beziehungen konnten in diesem Fall als gering bis mittel beurteilt werden.

*Zusammenhang des familiären Hintergrunds mit M und M'*. Die Indikatoren des familiären Hintergrunds korrelierten mittel (in Bezug auf den ISEI) bzw. mittel bis stark (in Bezug auf die Anzahl an Büchern) mit der generellen mathematischen Fähigkeit. Hingegen waren erwartungskonform die Zusammenhänge dieser Indikatoren mit der mathematikspezifischen Fähigkeit nahezu Null.

Inferenzstatistisch konnten die substanziellen Korrelationen der generellen mathematischen Fähigkeit nicht mit dem  $\chi^2$ -Differenzentest abgesichert werden, da für diesen Test bei beiden Indikatoren des SES negative Werte resultierten. Jedoch zeigt Tabelle 27 offensichtlich, dass sich der Modell-Fit verschlechterte. Ebenso ist in Abbildung 30b erkennbar, dass die 95%-Konfidenzintervalle dieser Korrelationen die Null nicht einschlossen.

Die Korrelation der mathematikspezifischen Fähigkeit mit der Anzahl an Büchern war erwartungskonform nicht statistisch signifikant von Null verschieden, hingegen aber die Korrelation mit dem ISEI. Zusammenfassend kann man damit festhalten, dass die Hypothesen 4c und 4d deskriptiv klar und inferenzstatistisch nur teilweise von den Daten gestützt wurden.

Die substanziellen positiven Korrelationen der generellen mathematischen Fähigkeit mit beiden Indikatoren des familiären Hintergrunds verringerten sich bei Kontrolle der Schulformzugehörigkeit deutlich und waren dann als gering bis mittel zu beurteilen. Hingegen blieben die korrespondierenden Korrelationen mit der mathematikspezifischen Fähigkeit weitestgehend unverändert.

Erwartungskonform waren die Zusammenhänge mit den Merkmalen des familiären Hintergrunds bei beiden mathematischen Fähigkeiten nicht gleich stark ausgeprägt. Die mathematikspezifische Fähigkeit korrelierte im Gegensatz zur generellen mathematischen Fähigkeit kaum mit den Indikatoren des familiären Hintergrunds. Die deskriptiv klar erkennbaren Unterschiede in den korrelativen Zusammenhängen (vgl. Abb. 30) waren auch statistisch signifikant verschieden. Hypothese 4e wurde somit deskriptiv wie auch inferenzstatistisch von den Daten gestützt.

#### 9.2.4 Diskussion

Welche zentralen Befunde zum Zusammenhang der kognitiven Fähigkeiten mathematischer Schülerleistung mit Indikatoren des familiären Hintergrunds können hier herausgestellt werden?

- Die allgemeine kognitive Fähigkeit und die fluide Fähigkeit korrelierten substanziell und mit nahezu gleicher Stärke mit den Indikatoren des familiären Hintergrunds (Hypothesen 4a und 4b). Bei Kontrolle der Schulformzugehörigkeit waren diese Zusammenhänge weniger stark.
- Mit Blick auf die mathematischen Fähigkeiten ist festzuhalten, dass hier deutliche Unterschiede in den Zusammenhängen resultierten (Hypothese 4e): Während die generelle mathematische Fähigkeit erwartungskonform positiv mit beiden Indikatoren korrelierte (Hypothese 4c), waren die korrespondierenden Zusammenhänge der mathematikspezifischen Fähigkeit Null (in Bezug auf die Anzahl an Büchern) oder nahezu Null (in Bezug auf den ISEI). Dieses letztgenannte Befundmuster stützte Hypothese 4d.
- Wenn die Schulformzugehörigkeit kontrolliert wurde, verringerten sich die Zusammenhänge zwischen den Merkmalen des familiären Hintergrunds mit der generellen mathematischen Fähigkeit und veränderten sich bei der mathematikspezifischen Fähigkeit nur unmerklich.

Die Ergebnisse zum Zusammenhang kognitiver Fähigkeiten mit Merkmalen des familiären Hintergrunds sind angesichts der Diskussion sozialer Disparitäten sehr aufschlussreich. Soziale Disparitäten (in Form von Unterschieden des familiären Hintergrunds) bestehen unabhängig von der Schulformzugehörigkeit, aber auch in Abhängigkeit von der Schulformzugehörigkeit *cum grano salis* nur in der allgemeinen kognitiven Fähigkeit (bzw. fluiden Fähigkeit). Schüler gleicher allgemeiner kognitiver Fähigkeit, aber mit unterschiedlichem sozioökonomischem oder kulturellem Kapital unterscheiden sich nicht bei der mathematikspezifischen Fähigkeit. Die Befunde der vorliegenden Arbeit deuten also darauf hin, dass soziale Disparitäten in der mathematischen Schülerleistung, wie sie in früheren Studien berichtet wurden (Baumert & Schümer, 2001, 2002; White, 1982), in erster Linie die Zusammenhänge zwischen der allgemeinen kognitiven Fähigkeit und den Merkmalen des familiären Hintergrunds eines Schülers widerspiegeln.

Weiterhin scheinen sich lernförderliche Unterschiede des familiären Hintergrunds im Sinne „allgemeiner Umweltunterschiede“ gleichermaßen auf eine Vielzahl kognitiver Fähigkeiten auszuwirken und nicht im Sinne von „verhaltensspezifischen Unterschieden“ nur auf die mathematischen Fähigkeiten (vgl. Abschnitt 4.3.1; Baltes, Nesselroade & Cornelius, 1978). Dies zeigt die substantiellen Korrelationen von Indikatoren des familiären Hintergrunds mit der allgemeinen kognitiven Fähigkeit und die äußerst geringen Korrelationen mit der mathematikspezifischen Fähigkeit. Dieser Befund stützt damit auch die Annahme, dass das Wissen und die Fähigkeiten, die zur Lösung von Aufgaben zur Messung mathematischer Schülerleistung benötigt werden, primär in der Schule erworben werden (vgl. Geary, 1995; Köller & Baumert, 2002). Vor diesem Hintergrund stellt sich für zukünftige Forschungsvorhaben die lohnende Frage, welchen Beitrag Unterschiede des Mathematikunterrichts am Zustandekommen der Unterschiede in der mathematikspezifischen Fähigkeit haben (siehe hierzu auch Abschnitt 10.3.2).

## 9.3 Selbstkonzept

### 9.3.1 Das Bezugsrahmenmodell von Marsh (1986)

Geschlecht und familiärer Hintergrund sind zwei soziodemografische Schülermerkmale, die häufig mit Unterschieden in kognitiven Fähigkeiten oder der Lernmotivation in Beziehung gesetzt werden. Als zwei zentrale Vertreter der Lernmotivation können im Rahmen von „Erwartung-Wert-Theorien“ das fachspezifische Selbstkonzept (als Erwartungskomponente) und das fachspezifische Interesse (als Wertkomponente) betrachtet werden (siehe Abschnitt 5.2.2 sowie z.B. Eccles & Wigfield, 2002; Pintrich, 2003).

In diesem Abschnitt steht das Selbstkonzept im Mittelpunkt (Abschnitt 9.4 befasst sich mit dem Interesse). Hierzu wird einerseits die *konvergente* Validität zwischen dem mathematischen Selbstkonzept und mathematischen Fähigkeiten untersucht. Andererseits wird auch die *diskriminante* Validität zum verbalen Selbstkonzept analysiert.

Hypothesen zur konvergenten und diskriminanten Validität des Selbstkonzepts können mithilfe des Bezugsrahmenmodells (*Internal/External Frame of Reference Model*, abgekürzt I/E-Modell) von Marsh (1986) abgeleitet werden (siehe Möller & Köller, 2004, für eine Zusammenfassung). Ausgangspunkt für die Entwicklung dieses Modells ist der Befund, dass mathematische und verbale Selbstkonzepte trotz hoher positiver Korrelationen zwischen den korrespondierenden Leistungsindikatoren nahezu unabhängig voneinander sind (Marsh, 1986). Dieser Befund wird im Bezugsrahmenmodell dadurch erklärt, dass zwei Informationsquellen genutzt werden, die zur Genese der Selbstkonzepte beitragen.

- (a) Bei (interindividuellen) *externalen Vergleichen* werden die eigenen Leistungen in einem Schulfach mit den Leistungen der Mitschüler verglichen (*external frame of reference*). Da die Leistungen in verschiedenen Schulfächern positiv interkorreliert sind, sollte dieser Vergleichsprozess dazu führen, dass das mathematische und verbale Selbstkonzept positiv miteinander korrelieren. Dies impliziert auch eine positive Korrelation zwischen Fachleistung und dem jeweiligen fachspezifischen Selbstkonzept.
- (b) Bei *internalen Vergleichen* werden die eigenen Leistungen in einem Fach mit den eigenen Leistungen in einem anderen Fach (intra-individuell) verglichen (*internal frame of reference*). Beim internalen Vergleich ist die Differenz zwischen den eigenen Fachleistungen entscheidend, un-

abhängig davon wie gut oder schlecht man im Vergleich zu den Mitschülern abschneidet. Sind zum Beispiel die mathematischen Leistungen eines Schülers größer als seine verbalen Leistungen (z.B. im Fach Deutsch), dann ist das mathematische Selbstkonzept dieses Schülers stärker ausgeprägt als sein verbales Selbstkonzept.

Die Annahmen des Bezugsrahmenmodells können gut mit einem pfadanalytischen Modell veranschaulicht werden (Abb. 27).

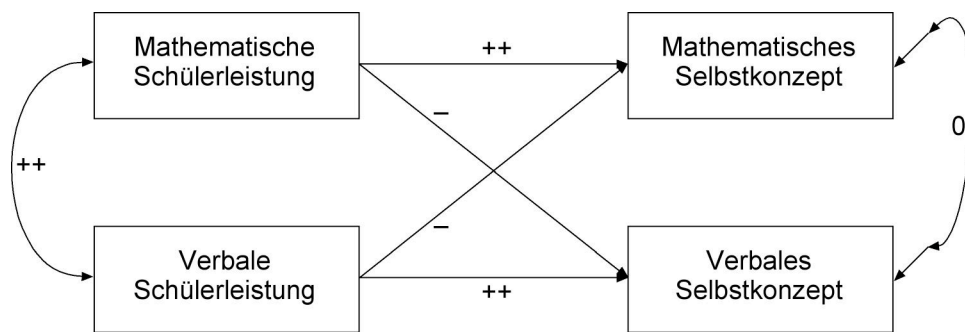


Abbildung 27: Bezugsrahmenmodell in Anlehnung an Marsh (1986, S. 134)

Die mathematische Schülerleistung und die verbale Schülerleistung sind positiv interkorreliert und haben jeweils positive Regressionsgewichte auf die korrespondierenden Selbstkonzepte. Dies entspricht der Wirkung des externalen Vergleichs.

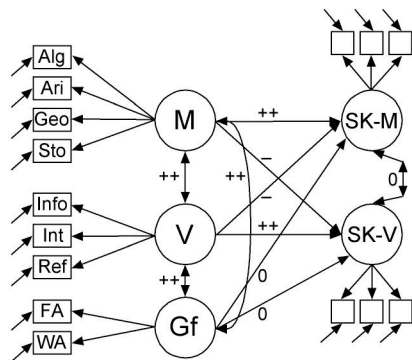
Hingegen hat die mathematische Schülerleistung ein negatives Regressionsgewicht auf das verbale Selbstkonzept, und die verbale Schülerleistung hat ein negatives Regressionsgewicht auf das mathematische Selbstkonzept. Dies entspricht der Wirkung des internalen Vergleichs. Zum Beispiel ist für Schüler mit der gleichen mathematischen Schülerleistung (die Fachleistung wird im Regressionsmodell konstant gehalten) ein stärkeres mathematisches Selbstkonzept und ein schwächeres verbales Selbstkonzept zu erwarten, wenn ihre mathematische Schülerleistung (beim intra-individuellen Vergleich) größer ist als ihre verbale Schülerleistung.

Nach Kontrolle der Fachleistungen interkorrelieren die Residualterme der fachspezifischen Selbstkonzepte nahezu Null. Der Grund hierfür ist, dass der soziale und dimensionale Vergleich gegenläufig operieren, was insgesamt zu einer „Netto“-Korrelation von nahezu Null führen kann.

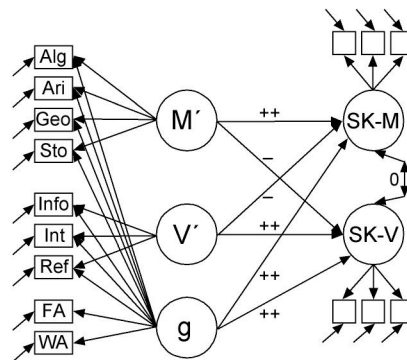
Wichtig ist, an dieser Stelle zu klären, wie die mathematische Schülerleistung und die verbale Schülerleistung aus dem Bezugsrahmenmodell von Marsh mit den kognitiven Fähigkeiten in dieser Arbeit zusammenhängen. Zur empirischen Prüfung des Bezugsrahmenmodells werden entweder Schulnoten (z.B. Rost, Dickhäuser, Sparfeldt & Schilling, 2004; Schilling, Sparfeldt & Rost, 2004) oder globale Leistungsscores aus standardisierten Schülerleistungstests verwendet (z.B. Lüdtke, Köller, Artelt, Stanat & Baumert, 2002; Marsh, 1986). In der vorliegenden Arbeit wird ein Forschungsansatz verfolgt, der Schülerleistungen aus der Perspektive kognitiver Fähigkeiten analysiert (vgl. Abschnitt 4.1). Daher werden hier Aufgaben zur Messung der mathematischen Schülerleistung als Indikatoren mathematischer Fähigkeiten (z.B. der generellen mathematischen Fähigkeit M) und Aufgaben zur Messung der verbalen Schülerleistung als Indikatoren verbaler Fähigkeiten (z.B. der generellen verbalen Fähigkeit V) betrachtet. Folgt man dieser Argumentation, dann beruht die empirische Prüfung des Bezugsrahmenmodells in Studien, die

Scores standardisierter Schülerleistungstests verwenden, auf dem Standardmodell. Dies ist in Abbildung 28a dargestellt. In den Abbildungen 28a und 28b sind im Gegensatz zur ursprünglichen Formulierung des Bezugsrahmenmodells von Marsh (1986) auch die fluide Fähigkeit und die allgemeine kognitive Fähigkeit eingezeichnet. Nachfolgend werden nun Annahmen formuliert, wie die jeweiligen kognitiven Fähigkeiten innerhalb der beiden Modelle mit dem mathematischen und verbalen Selbstkonzept assoziiert sind.

(a) Konzeption in Form des Standardmodells



(b) Konzeption in Form des Nested-Faktormodells



M = generelle mathematische Fähigkeit, V = generelle verbale Fähigkeit, Gf = fluide Fähigkeit, M' = mathematikspezifische Fähigkeit, V' = spezifische verbale Fähigkeit, g = allgemeine kognitive Fähigkeit, SK-M = mathematisches Selbstkonzept, SK-V = verbales Selbstkonzept. Für eine Beschreibung der manifesten Variablen siehe Abschnitt 7.2.2.

Abbildung 28: Erwarteter Effekt für das Bezugsrahmenmodell

*Standardmodell.* Folgt man dem Bezugsrahmenmodell, dann korrelieren höhere Fachleistungen mit höheren fachspezifischen Selbstkonzepten. Da die fluide Fähigkeit entsprechend den Modellen schulischen Lernens ein bedeutsamer Prädiktor der mathematischen sowie der verbalen Schülerleistung ist (vgl. Abschnitt 5.1), ist zu erwarten, dass die fluide Fähigkeit positiv mit dem mathematischen und dem verbalen Selbstkonzept korreliert.

Jedoch ist im pfadanalytischen Modell bei der Regression der beiden Selbstkonzepte auf die fluide Fähigkeit ein Effekt um Null zu erwarten. Der Grund hierfür wird beispielhaft anhand der generellen mathematischen Fähigkeit und dem mathematischen Selbstkonzept erläutert (die Argumentation verläuft analog für die generelle verbale Fähigkeit). Es ist anzunehmen, dass das mathematische Selbstkonzept stärker mit der generellen mathematischen Fähigkeit korreliert als die fluide Fähigkeit. In die generelle mathematische Fähigkeit gehen entsprechend den Annahmen des Nested-Faktormodells die allgemeine kognitive Fähigkeit und die mathematikspezifische Fähigkeit gemischt ein, und die letzten beiden Fähigkeiten korrelieren jeweils positiv mit dem mathematischen Selbstkonzept. Bei einer Regression des mathematischen Selbstkonzepts auf mehrere Prädiktoren wird dann in der Regel der Prädiktor am stärksten gewichtet, der am höchsten mit dem jeweiligen Kriterium korreliert (Cohen & Cohen, 1983). Dies ist in diesem Fall also die generelle mathematische Fähigkeit. Aufgrund der Kolinearität (bzw. der Interkorrelation) der fluiden Fähigkeit mit der generellen mathematischen Fähigkeit wird dann der gemeinsame Varianzanteil von der fluiden Fähigkeit und dem mathematischen Selbstkonzept durch die generelle mathematische Fähigkeit erklärt. Das Regressionsgewicht der fluiden Fähigkeit repräsentiert in diesem Fall den inkrementellen Varianzanteil, den sie über die generelle mathematische Fähigkeit

hinaus am mathematischen Selbstkonzept hat. Dieser dürfte (auch aufgrund der hohen Interkorrelation von .88 der fluiden Fähigkeit und der generellen mathematischen Fähigkeit in der vorliegenden Arbeit, siehe Abschnitt 7.2.3) um Null sein.

Vor dem Hintergrund der obigen Überlegung ist auch davon auszugehen, dass – wie in der ursprünglichen Formulierung des Bezugsrahmenmodells vorgesehen – die generelle mathematische Fähigkeit einen positiven Effekt auf das mathematische Selbstkonzept und einen negativen Effekt auf das verbale Selbstkonzept hat. Hingegen sollte die generelle verbale Fähigkeit einen positiven Effekt auf das verbale Selbstkonzept und einen negativen Effekt auf das mathematische Selbstkonzept aufweisen.

*Nested-Faktormodell.* Im Gegensatz zur fluiden Fähigkeit wird für die allgemeine kognitive Fähigkeit ein positiver Effekt auf das mathematische und das verbale Selbstkonzept erwartet. Hierfür spricht, dass die allgemeine kognitive Fähigkeit jeweils mit höheren Fachleistungen und folglich mit höheren fachspezifischen Selbstkonzepten einhergehen sollte. Weiterhin ist sie unkorreliert mit den fachspezifischen Fähigkeiten. Damit besteht nicht das oben beschriebene Problem der Kolinearität.

Mit Blick auf die mathematikspezifische Fähigkeit ist zu erwarten, dass diese mit höheren Fachleistungen einhergeht und somit auch einen positiven Effekt auf das mathematische Selbstkonzept haben sollte. Ebenso sollten höhere Leistungen bei der spezifischen verbalen Fähigkeit mit einem höheren verbalen Selbstkonzept assoziiert sein.

Es stellt sich nun die Frage, welchen Effekt die mathematikspezifische Fähigkeit auf das verbale Selbstkonzept hat. Eine regressionsanalytische Kontrolle für die spezifische verbale Fähigkeit und die allgemeine kognitive Fähigkeit im verbalen Selbstkonzept ist weitestgehend äquivalent zur Kontrolle für die generelle verbale Fähigkeit. Entsprechend den Annahmen des Bezugsrahmenmodells sollte daher der Effekt der mathematikspezifischen Fähigkeit auf das verbale Selbstkonzept negativ sein. Die gleiche Überlegung greift für die spezifische verbale Fähigkeit: Diese sollte einen negativen Effekt auf das mathematische Selbstkonzept aufweisen.

### 9.3.2 Methode

*Manifeste Variablen.* Indikatoren der fachspezifischen Selbstkonzepte (siehe Tab. 28) waren jeweils drei Items aus dem *Self Description Questionnaire* (SDQ, Marsh, 1990), die bei PISA 2000 eingesetzt wurden (Kunter u.a., 2002). Schüler konnten bei allen Items mit einer vierstufigen Skala (1 = trifft nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu, 4 = trifft zu) Stellung zu den Aussagen nehmen, die ihr mathematisches oder verbales Selbstkonzept thematisierten.

*Analyseverfahren.* Das Bezugsrahmenmodell wurde jeweils für die Konzeption der kognitiven Fähigkeiten mathematischer Schülerleistung im Rahmen des Standardmodells und des Nested-Faktormodells untersucht. Die latenten Variablen für die jeweiligen kognitiven Fähigkeiten wurden spezifiziert, wie es in Abschnitt 7.2.2 beschrieben wurde. Die latenten Variablen für die fachspezifischen Selbstkonzepte wurden identifiziert, indem jeweils die Faktorladung eines Indikators auf 1,0 fixiert wurde (siehe Tab. 28). Alle anderen Modellparameter wurden frei mit dem Programm Mplus 3.01 (Muthén & Muthén, 1998–2004b) bei Verwendung des Moduls *complex* und dem Schätzalgorithmus MLR geschätzt. Dieses Vorgehen hatte den Vorteil, dass so der Modell-Fit für die Messmodelle für die kognitiven Fähigkeiten und die motivationalen Konstrukte sowie die Beziehungen zwischen den latenten Variablen simultan analysiert werden konnten.

Tabelle 28: Items zur Messung des mathematischen und verbalen Selbstkonzepts

Itemtext	M	SD	#fehlend
<i>Mathematisches Selbstkonzept</i> ( $\omega = .89$ )			
Im Fach Mathematik bekomme ich gute Noten <sup>1</sup>	2,65	0,98	1.478
Mathematik ist eines meiner besten Fächer	2,41	1,09	1.595
Ich war schon immer gut in Mathematik	2,39	1,02	1.591
<i>Verbales Selbstkonzept</i> ( $\omega = .81$ )			
Im Fach Deutsch bin ich ein hoffnungsloser Fall <sup>1, 2</sup>	1,91	0,89	1.163
Im Fach Deutsch lerne ich schnell	2,73	0,87	1.482
Im Fach Deutsch bekomme ich gute Noten	2,77	0,87	1.590

Mittelwerte und Standardabweichungen beziehen sich auf die mit dem *Full Information Maximum Likelihood* (FIML) Schätzer ermittelten Stichprobenkennwerte. #fehlend = absolute Anzahl fehlender Werte.  $\omega$  = Skalenreliabilität berechnet nach McDonald (1999, S. 89) auf Grundlage der mit FIML geschätzten unstandardisierten Faktorladungen und unstandardisierten Residualvarianzen der manifesten Variablen (siehe Enders, 2003, der eine vergleichbare Prozedur für EM-imputierte Werte und Cronbachs Alpha vorschlägt).

<sup>1</sup> Faktorladung des Items wurde auf 1,0 fixiert, um den Faktor zu identifizieren.

<sup>2</sup> Item wurde bei der Reliabilitätsberechnung umgepolt.

Zusätzlich zu einer regressionsanalytischen Spezifikation (vgl. Abb. 28) wurden auch die korrelativen Beziehungen zwischen den kognitiven Fähigkeiten im Standardmodell bzw. dem Nested-Faktormodell und den jeweiligen fachspezifischen Selbstkonzepten analysiert. Hierbei wurden – wie oben beschrieben – die jeweiligen Modellparameter frei geschätzt. Die so ermittelten latenten Korrelationen sind in den Tabellen 30 und 31 enthalten.

Darüber hinaus sind in Abbildung 30 die latenten Korrelationen eingezeichnet, die nach der in Abschnitt 9.1.2 beschriebenen Methodik ermittelt wurden.

Weiterhin wurde das Bezugsrahmenmodell für das Standardmodell wie auch das Nested-Faktormodell auch bei Kontrolle der Schulformzugehörigkeit untersucht (siehe Abschnitt 9.1.2 für die Beschreibung der Methodik), um für mögliche Effekte der Schulform zu kontrollieren.

### 9.3.3 Ergebnisse

*Evaluation des Modell-Fits.* Der Modell-Fit für das Standardmodell wie auch für das Nested-Faktormodell war als gut zu bewerten (Tab. 29). Dies stützte die weitere Interpretation der Modellparameter.

*Latente Korrelationen.* Bevor auf die Regressionskoeffizienten eingegangen wird, sollen zunächst die latenten Korrelationen betrachtet werden. In Tabelle 30 sind die latenten Korrelationen für das Standardmodell und in Tabelle 31 sind die korrespondierenden Korrelationen für das Nested-Faktormodell eingetragen.



Tabelle 29: Globale Modell-Fit-Indizes der Strukturmodelle zur Analyse des Bezugsrahmenmodells

Modell	$\chi^2$	df	SF	CFI	RMSEA	SRMR
<i>Ohne Kontrolle der Schulform</i>						
Standardmodell	3.552	81	1,066	.99	.04	.03
Nested-Faktormodell	2.799	76	1,066	.99	.04	.03
<i>Bei Kontrolle der Schulform</i>						
Standardmodell	2.852	81	1,063	.98	.03	.03
Nested-Faktormodell	2.640	76	1,060	.98	.03	.03

SF = Skalierungsfaktor.

Tabelle 30: Standardmodell: Latente Korrelationen der kognitiven Fähigkeiten mit den fachspezifischen Selbstkonzepten

	M	V	Gf	SK-V	SK-M
M		.64	.79	-.03	.46
V	.80		.70	.23	.10
Gf	.89	.83		.06	.30
SK-V	.07	.25	.13		-.13
SK-M	.34	.06	.22	-.14	

Der Modell-Fit für das Modell zur Berechnung der Korrelationen war identisch mit den Werten, die in Tabelle 29 für das Standardmodell berichtet wurden. Unterhalb der Diagonale sind die Korrelationen ohne Kontrolle der Schulformzugehörigkeit eingetragen. Über der Diagonale stehen die latenten Korrelationen bei Kontrolle der Schulformzugehörigkeit. M = generelle mathematische Fähigkeit, V = generelle verbale Fähigkeit, Gf = fluide Fähigkeit, SK-V = verbales Selbstkonzept, SK-M = mathematisches Selbstkonzept.

Tabelle 31: Nested-Faktormodell: Latente Korrelationen der kognitiven Fähigkeiten mit den fachspezifischen Selbstkonzepten

	M'	V'	g	SK-V	SK-M
M'		.00 <sup>a</sup>	.00 <sup>a</sup>	-.17	.39
V'	.00 <sup>a</sup>		.00 <sup>a</sup>	.27	-.20
g	.00 <sup>a</sup>	.00 <sup>a</sup>		.07	.30
SK-V	-.17	.26	.15		-.14
SK-M	.36	-.24	.21	-.14	

Der Modell-Fit für das Modell zur Berechnung der Korrelationen war identisch mit den Werten, die in Tabelle 29 für das Nested-Faktormodell berichtet wurden. Unterhalb der Diagonale sind die Korrelationen ohne Kontrolle der Schulformzugehörigkeit eingetragen. Über der Diagonale stehen die latenten Korrelationen bei Kontrolle der Schulformzugehörigkeit. M' = mathematikspezifische Fähigkeit, V' = spezifische verbale Fähigkeit, g = allgemeine kognitive Fähigkeit, SK-V = verbales Selbstkonzept, SK-M = mathematisches Selbstkonzept.

<sup>a</sup> Diese Korrelationen wurden auf Null fixiert.

Mit Blick auf die latenten Korrelationen sind mehrere Aspekte hervorzuheben:

- Auch bei Kontrolle der Schulformzugehörigkeit blieb in beiden Modellen das Muster der Interkorrelationen insgesamt gesehen weitestgehend unverändert. Die Veränderungen der Korrelationen konnten (unter anderem) auf die Varianzrestriktionen von bis zu 28 Prozent in den manifesten Variablen zur Messung der kognitiven Fähigkeit attribuiert werden, die mit einer Kontrolle der Schulformzugehörigkeit einhergingen (siehe Anhang).
- Die fluide Fähigkeit wie auch die allgemeine kognitive Fähigkeit korrelierten jeweils positiv und in nahezu gleicher Höhe mit dem verbalen und mathematischen Selbstkonzept. Die Höhe der Korrelation war – entsprechend der Terminologie von Cohen (1992) – als gering bis mittel zu beurteilen.
- Die mathematikspezifische und die spezifische verbale Fähigkeit korrelierten jeweils positiv (in mittlerer Höhe) mit den korrespondierenden fachspezifischen Selbstkonzepten und jeweils negativ (in mittlerer Höhe) mit den Selbstkonzepten der jeweils anderen Domäne. Mit Blick auf dieses letzte Resultat zeigte sich ein deutlicher Unterschied zu den korrespondierenden Fähigkeiten im Standardmodell, die jeweils geringfügig positiv mit den Selbstkonzepten der anderen Inhaltsdomäne korrelierten.

*Analysen des Bezugsrahmenmodells.* Diese korrelativen Befunde stützten auch die Annahmen der zu erwartenden Effekte auf die fachspezifischen Selbstkonzepte, die bei einer Konzeptualisierung der kognitiven Fähigkeiten im Rahmen des Nested-Faktormodells gemacht wurden. Denn bei der pfadanalytischen Auswertung des Bezugsrahmenmodells hatten die mathematikspezifische und die spezifische verbale Fähigkeit jeweils positive Effekte auf das Selbstkonzept der korrespondierenden Inhaltsdomäne und einen negativen Effekt auf das Selbstkonzept der jeweils anderen Inhaltsdomäne (Tab. 32). Weiterhin hatte die allgemeine kognitive Fähigkeit einen positiven Effekt auf beide Selbstkonzepte.

Tabelle 32: Analysen des Bezugsrahmenmodells für das Standardmodell und des Nested-Faktormodells: standardisierte Modellparameter (in Klammern: Modellparameter bei Kontrolle der Schulformzugehörigkeit)

Unabhängige Variable	Standardmodell		Nested-Faktormodell	
	Abhängige Variable SK-M	Abhängige Variable SK-V	Unabhängige Variable	Abhängige Variable SK-M SK-V
M	.81* (.68)	-.39* (-.31)	M'	.36* (.39) -.17* (-.17)
V	-.55* (-.34)	.55* (.41)	V'	-.24* (-.20) .26* (.27)
Gf	-.05 (-.01)	.03 (.02)	g	.21* (.30) .15* (.07)
R <sup>2</sup>	.23 (.28)	.12 (.11)	R <sup>2</sup>	.23 (.28) .12 (.11)

\*  $p < .05$  (einseitiger Test).

M = generelle mathematische Fähigkeit, V = generelle verbale Fähigkeit, Gf = fluide Fähigkeit, M' = mathematikspezifische Fähigkeit, V' = spezifische verbale Fähigkeit, g = allgemeine kognitive Fähigkeit, SK-M = mathematisches Selbstkonzept, SK-V = verbales Selbstkonzept. Die Korrelation zwischen den Residuen von SK-M und SK-V betrug im Standardmodell und im Nested-Faktormodell jeweils  $-.05$ .

Erwartungskonform hatte im Standardmodell die fluide Fähigkeit – trotz der positiven Korrelationen mit den beiden fachspezifischen Selbstkonzepten – nahezu keinen Einfluss auf die beiden Selbstkonzepte, wenn für die beiden anderen kognitiven Fähigkeiten kontrolliert wurde. Für die generelle mathematische und die generelle verbale Fähigkeit resultierte das angenommene Befundmuster, das analog zu den inhaltspezifischen Fähigkeiten im Nested-Faktormodell war. Im Vergleich hierzu waren lediglich die Regressionskoeffizienten deutlich stärker.

Insgesamt gesehen war in beiden Modellen das Ausmaß der an den fachspezifischen Selbstkonzepten aufgeklärten Varianz ( $R^2$ ) identisch.

Ein analoges Befundmuster zeigte sich für beide Modelle auch bei Kontrolle der Schulform. Die Unterschiede zwischen den standardisierten Modellparametern bei den Modellen mit und ohne Kontrolle der Schulformzugehörigkeit sind wohl in erster Linie auf die Varianzrestriktionen und den damit verbundenen Implikationen bei der Berechnung standardisierter Parameter zurückzuführen.

### 9.3.4 Diskussion

Welche zentralen Befunde zum Zusammenhang der kognitiven Fähigkeiten mit den fachspezifischen Selbstkonzepten konnten gewonnen werden?

- Die anhand des Bezugsrahmenmodells abgeleiteten Beziehungen konnten allesamt bestätigt werden.
- Die allgemeine kognitive Fähigkeit wie auch die fluide Fähigkeit korrelierten jeweils positiv mit dem mathematischen Selbstkonzept wie auch mit dem verbalen Selbstkonzept. Die korrespondierenden Korrelationen waren fast identisch.
- Bei der regressionsanalytischen Auswertung des Bezugsrahmenmodells würde (ohne Betrachtung der bivariaten Korrelationen) der Zusammenhang der fluiden Fähigkeit mit den fachspezifischen Selbstkonzepten unterschätzt werden. Dies war bei Verwendung des Nested-Faktormodells nicht der Fall.
- Die mathematikspezifische Fähigkeit hatte einen positiven Effekt auf das mathematische Selbstkonzept und einen negativen Effekt auf das verbale Selbstkonzept. Dieses Befundmuster korrespondierte auch mit den korrelativen Beziehungen der mathematikspezifischen Fähigkeit mit den beiden Selbstkonzepten.
- Im Gegensatz hierzu korrelierten die generelle mathematische und die generelle verbale Fähigkeit jeweils geringfügig positiv mit den Selbstkonzepten der jeweils anderen Inhaltsdomäne.

Das korrelative Befundmuster machte insgesamt deutlich, dass bei alleiniger Betrachtung der generellen mathematischen Fähigkeit (ohne die regressionsanalytische Kontrolle für die generelle verbale Fähigkeit und die fluide Fähigkeit) Beziehungsgeflechte zu wichtigen psychologischen Variablen – in diesem Fall den fachspezifischen Selbstkonzepten – weniger detailliert aufgelöst werden konnten. Der Grund hierfür ist, dass in die generelle mathematische Fähigkeit die allgemeine kognitive Fähigkeit und die mathematikspezifische Fähigkeit gemischt eingingen. Insbe-

sondere mit Blick auf die diskriminante Validität zum verbalen Selbstkonzept wurde dies aufgrund der differenziellen Validität der mathematikspezifischen Fähigkeit und der allgemeinen kognitiven Fähigkeit offensichtlich.

Aufschlussreich war nun, dass sich bei der pfadanalytischen Untersuchung zum erweiterten Bezugsrahmenmodell für das Standardmodell und das Nested-Faktormodell äquivalente Befunde für die spezifischen Fähigkeiten ( $M'$  und  $V'$ ) und die generellen Fähigkeiten ( $M$  und  $V$ ) ergaben. So waren die Effekte zu korrespondierenden domänenspezifischen Selbstkonzepten jeweils positiv. Domänenübergreifende Effekte waren jeweils negativ. Warum das so ist, soll nachfolgend kurz erläutert werden. Im Standardmodell konnten die fluide Fähigkeit, die generelle verbale Fähigkeit und die generelle mathematische Fähigkeit die Selbstkonzepte beeinflussen. Dadurch wurde auch für den Varianzanteil in der abhängigen Variablen kontrolliert, der auf die gemeinsame Varianz zwischen den drei kognitiven Fähigkeiten zurückging. Dieser Anteil sollte (unter der Äquivalenzannahme) dem Varianzanteil der allgemeinen kognitiven Fähigkeit in der abhängigen Variablen entsprechen. Im Nested-Faktormodell sind die spezifischen Fähigkeiten und die allgemeine kognitive Fähigkeit orthogonal. Damit ist es aufgrund der spezifizierten Konzeptualisierung der kognitiven Fähigkeiten unmöglich, dass diese drei kognitiven Fähigkeiten gemeinsame Varianzanteile in den abhängigen Variablen aufweisen. Das heißt, bei der pfadanalytischen Untersuchung des Bezugsrahmenmodells wurde im Nested-Faktormodell a priori für die gemeinsame Varianz in den abhängigen Variablen kontrolliert, wohingegen im Standardmodell erst durch das Regressionsmodell für diese gemeinsame Varianz kontrolliert wurde.

Weiterhin zeigten die Ergebnisse die theoretische Stärke des Bezugsrahmenmodells auf. Alle Hypothesen konnten anhand der Modellannahmen oder -implikationen abgeleitet werden. Die vorliegende Arbeit leistet aber auch einen Beitrag, der die Weiterentwicklung des Modells stimulieren könnte. In bisherigen Analysen wurde die allgemeine kognitive Fähigkeit nicht berücksichtigt, um das Bezugsrahmenmodell zu untersuchen. Mit der vorliegenden Arbeit und dem Ansatz eines *coordinate measurement* von Messick (1984) wurde ein erster Schritt gemacht, eines der „klassischen“ Konstrukte aus der Strukturforschung kognitiver Fähigkeiten in dieses Modell zu integrieren.

Folgt man diesem Ansatz, dann öffnet sich eine interessante Interpretationsmöglichkeit, welche kognitiven Fähigkeiten bei den jeweiligen Vergleichsprozessen eine Rolle spielen. Bei den externalen Vergleichsprozessen, die bei der Bildung des mathematischen Selbstkonzepts beteiligt sind, operieren die allgemeine kognitive Fähigkeit und die mathematikspezifische Fähigkeit (bzw. die spezifische verbale Fähigkeit) gleichsinnig. Hingegen weisen sie in entgegengesetzte Richtungen mit Blick auf das verbale (bzw. das mathematische) Selbstkonzept. Basiert der interne Vergleich somit auf einem Vergleich der mathematikspezifischen und der spezifischen verbalen Fähigkeit? Diese Annahme wurde bisher nicht geprüft. Die Analysen in dieser Arbeit ermuntern also, über die theoretische Weiterentwicklung des Bezugsrahmenmodells hinsichtlich der Prädiktoren der fachspezifischen Selbstkonzepte nachzudenken.

## 9.4 Interesse

### 9.4.1 Ein erweitertes Bezugsrahmenmodell

Das fachspezifische Selbstkonzept repräsentiert in dieser Arbeit die Erwartungskomponente der „Erwartung-Wert-Theorien“ (Eccles & Wigfield, 2002; Pintrich, 2003). Dieser Abschnitt befasst

sich mit einem Vertreter der Wertkomponente: dem fachspezifischen Interesse. Zur konvergen-ten Validierung wird der Zusammenhang der kognitiven Fähigkeiten mathematischer Schülerleistung mit dem mathematischen Interesse (INT-M) untersucht. Die diskriminante Validität wird anhand des Interesses am Lesen (INT-L) analysiert.

Im Theorieteil wurde in Abschnitt 5.2.2 die mathematische Lernmotivation – entsprechend den Modellen schulischen Lernens – als Prädiktor des Lernerfolgs in Mathematik betrachtet. Allerdings sprechen einige Argumente auch dafür, dass die erzielte Leistung einen positiven Effekt auf das fachspezifische Interesse hat. Diese Wirkrichtung wird in diesem Abschnitt genutzt, um Hypothesen zum Zusammenhang zwischen kognitiven Fähigkeiten und dem mathematischen Interesse sowie dem Interesse am Lesen abzuleiten.

Entscheidend für die Wirkung der erzielten Leistung auf das Interesse ist das fachspezifische Kompetenzerleben. Eccles und Kollegen (z.B. Wigfield & Eccles, 2000) nehmen an, dass Kompetenz- und Fähigkeitseinschätzungen (z.B. aufgrund von Rückmeldeprozessen der eigenen Leistung) zentral für die Interessensgenese sind.

Als Maß für das Kompetenzerleben kann in Anlehnung an Köller und Kollegen (Köller, Daniels & Baumert, 2000; Köller, Schnabel & Baumert, 2000; Nagy, Trautwein, Baumert, Köller & Garrett, in Druck) das fachspezifische Selbstkonzept betrachtet werden. In den referierten Arbeiten wird angenommen, dass die fachspezifischen Selbstkonzepte den fachspezifischen Interessen kausal vorgeschaltet sind. Höhere fachspezifische Selbstkonzepte gehen mit einem höheren (korrespondierenden) fachspezifischen Interesse einher.

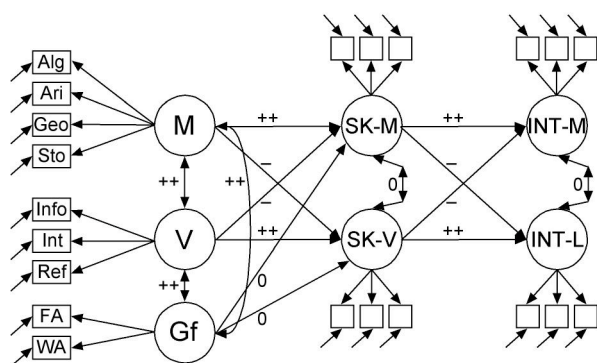
Weiterhin wurde in zwei Arbeiten von Köller und Kollegen der wechselseitige Einfluss der Fachleistungen, Selbstkonzepte und des Interesses für die Fächer Mathematik und Englisch (Köller, Daniels u.a., 2000) und für die Fächer Mathematik und Biologie (Nagy u.a., in Druck) untersucht. Hierzu erweiterten Köller und Kollegen das Bezugsrahmenmodell, indem sie auch reziproke Beziehungen zwischen den fachspezifischen Selbstkonzepten und den fachspezifischen Interessen annahmen: Das Selbstkonzept eines Fachs hatte analog zu den Annahmen im Bezugsrahmenmodell einen negativen Effekt auf das Interesse des anderen Fachs.<sup>29</sup> Zum Beispiel wirkte in der Studie von Nagy u.a. das Selbstkonzept in Biologie negativ auf das mathematische Interesse. Dabei wird angenommen, dass analog zum Bezugsrahmenmodell internale und externale Vergleichsprozesse die Genese von Selbstkonzepten und die Genese von Interessen beeinflussen (Nagy u.a., in Druck).

Folgt man diesen Annahmen, dann sind die folgenden Beziehungen für die kognitiven Fähigkeiten, Selbstkonzepte und Interessen im Rahmen eines *erweiterten Bezugsrahmenmodells* zu erwarten (Abb. 29). Dieses Modell wird sowohl für die Konzeptualisierung der kognitiven Fähigkeiten der mathematischen Schülerleistung in Form des Standardmodells (Abb. 29a) als auch in Form des Nested-Faktormodells (Abb. 29b) geprüft werden.

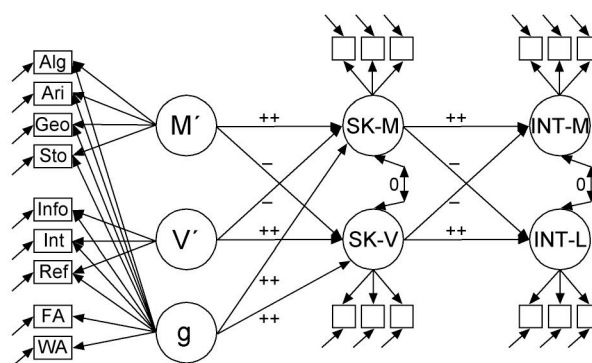
---

<sup>29</sup> Hierbei ist anzumerken, dass dieser Effekt in der Arbeit von Nagy und Kollegen nur für die Jungen festgestellt wurde. Für Mädchen resultierten jeweils Regressionsgewichte um Null vom mathematischen Selbstkonzept auf das Interesse in Biologie bzw. vom Selbstkonzept in Biologie auf das mathematische Interesse. Allerdings war der jeweilige Effekt in der Gesamtstichprobe von Jungen *und* Mädchen insgesamt leicht negativ.

(a) Standardmodell



(b) Nested-Faktormodell



M = generelle mathematische Fähigkeit, V = generelle verbale Fähigkeit, Gf = fluide Fähigkeit, SK-M = mathematisches Selbstkonzept, SK-V = verbales Selbstkonzept, INT-M = mathematisches Interesse, INT-L = Interesse am Lesen, M' = mathematikspezifische Fähigkeit, V' = spezifische verbale Fähigkeit, g = allgemeine kognitive Fähigkeit.

Abbildung 29: Erweitertes Bezugsrahmenmodell in Anlehnung an Nagy u.a. (in Druck)

*Standardmodell.* Die generelle mathematische Fähigkeit hat einen positiven Einfluss auf das mathematische Selbstkonzept. Dieses wiederum beeinflusst das mathematische Interesse positiv. Der gleiche Zusammenhang sollte für die generelle verbale Fähigkeit, das verbale Selbstkonzept und das Interesse am Lesen gelten. Hingegen wird angenommen, dass die fächer- bzw. domänenübergreifenden Pfade jeweils negativ sind. Darüber hinaus hat die fluide Fähigkeit keinen Effekt auf die fachspezifischen Selbstkonzepte (siehe hierzu die Erklärung in Abschnitt 9.3.1).

*Nested-Faktormodell.* Die allgemeine kognitive Fähigkeit hat, wie in Abschnitt 9.3 zum Bezugsrahmenmodell argumentiert wurde, einen positiven Effekt auf das mathematische und das verbale Selbstkonzept. Die mathematikspezifische Fähigkeit hat einen positiven Einfluss auf das mathematische Selbstkonzept. Das mathematische Selbstkonzept wiederum beeinflusst das mathematische Interesse positiv. Der gleiche Zusammenhang sollte für die spezifische verbale Fähigkeit, das verbale Selbstkonzept und das Interesse am Lesen gelten. Hingegen wird angenommen, dass domänenübergreifende Pfade zwischen Fähigkeiten und Selbstkonzepten bzw. Selbstkonzepten und Interessen jeweils negativ sind.

*Explorative Modelle.* In einem weiteren exploratorischen Schritt wird über die Prüfung des erweiterten Bezugsgruppenmodells hinaus noch der Frage nachgegangen, ob die kognitiven Fähigkeiten einen direkten Effekt auf die fachspezifischen Interessen haben. Deshalb wurden ein Standardmodell (*Standardmodell-direkt*) und ein Nested-Faktormodell (*Nested-Faktormodell-direkt*) spezifiziert, in denen die kognitiven Fähigkeiten einerseits jeweils die beiden Selbstkonzepte und andererseits jeweils (zusätzlich zu den Selbstkonzepten) die beiden fachspezifischen Interessen beeinflussten.

## 9.4.2 Methode

*Manifeste Variablen.* Indikatoren der beiden Interessenskonstrukte (siehe Tab. 33) waren jeweils drei Items, die bei PISA 2000 eingesetzt wurden (Kunter u.a., 2002). Schüler konnten bei allen Items mit einer vierstufigen Skala (1 = trifft nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu, 4 = trifft zu) Stellung zu den jeweiligen Aussagen nehmen.

Tabelle 33: Items zur Messung des mathematischen Interesses und des Interesses am Lesen

Itemtext	M	SD	#fehlend
<i>Mathematisches Interesse (<math>\omega = .77</math>)</i>			
Wenn ich mich mit Mathematik beschäftige, vergesse ich manchmal alles um mich herum <sup>1</sup>	2,25	0,92	1.194
Mathematik ist mir persönlich wichtig	2,57	1,01	1.640
Weil mir die Beschäftigung mit Mathematik Spaß macht, würde ich das nicht gerne aufgeben	2,33	1,02	1.298
<i>Interesse am Lesen (<math>\omega = .84</math>)</i>			
Weil mir das Lesen Spaß macht, würde ich es nicht gerne aufgeben <sup>1</sup>	2,43	1,08	1.382
Ich lese in meiner Freizeit	2,52	1,14	1.508
Wenn ich lese, vergesse ich manchmal alles um mich herum	2,52	1,13	1.536

Mittelwerte und Standardabweichungen beziehen sich auf die mit dem *Full Information Maximum Likelihood* (FIML) Schätzer ermittelten Stichprobenkennwerte. #fehlend = absolute Anzahl fehlender Werte,  $\omega$  = Skalenreliabilität berechnet nach McDonald (1999, S. 89) auf Grundlage der mit FIML geschätzten unstandardisierten Faktorladungen und unstandardisierten Residualvarianzen der manifesten Variablen (siehe Enders, 2003, der eine vergleichbare Prozedur für EM-imputierte Werte und Cronbachs Alpha vorschlägt).

<sup>1</sup> Faktorladung des Items wurde auf 1,0 fixiert, um den Faktor zu identifizieren.

*Analyseverfahren.* Das erweiterte Bezugsrahmenmodell wurde jeweils für die Konzeption der kognitiven Fähigkeiten mathematischer Schülerleistung im Rahmen des Standardmodells und des Nested-Faktormodells untersucht. Die Spezifikation der latenten Variablen, die die kognitiven Fähigkeiten und die Selbstkonzepte repräsentierten, war identisch mit der Spezifikation zur Analyse des Bezugsrahmenmodells (siehe Abschnitt 9.3). Die fachspezifischen Interessen wurden identifiziert, indem jeweils die Faktorladung eines Indikators auf 1,0 fixiert wurde (siehe Tab. 33). Alle anderen Modellparameter wurden frei mit dem Programm Mplus 3.01 (Muthén & Muthén, 1998–2004b) bei Verwendung des Moduls *complex* und des Schätzalgorithmus MLR geschätzt.

Auf Grundlage der so spezifizierten Messmodelle wurden zusätzlich zu einer regressionsanalytischen Spezifikation (vgl. Abb. 29a und 29b) auch die korrelativen Beziehungen zwischen den kognitiven Fähigkeiten im Standardmodell bzw. dem Nested-Faktormodell und den jeweiligen fachspezifischen Selbstkonzepten sowie fachspezifischen Interessen analysiert. Ferner sind in Abbildung 30 die latenten Korrelationen eingezeichnet, die nach der in Abschnitt 9.1.2 beschriebenen Methodik ermittelt wurden.

Um für mögliche Effekte der Schulform zu kontrollieren, wurden die Analysen auch bei Kontrolle der Schulformzugehörigkeit durchgeführt (siehe Abschnitt 9.1.2 für die Beschreibung der Methodik).

### 9.4.3 Ergebnisse

*Evaluation des Modell-Fits.* Der Modell-Fit aller spezifizierten Modelle war als gut zu bewerten (Tab. 34). Dies stützte die weitere Interpretation der Modellparameter. Allerdings hatten beide

Modelle, in denen direkte Effekte der kognitiven Fähigkeiten auf die fachspezifischen Interessen zulässig waren, einen besseren Modell-Fit. Dieses Ergebnis wird bei der Darstellung der Regressionskoeffizienten nochmals aufgegriffen.

Tabelle 34: Globale Modell-Fit-Indizes der Strukturmodelle zur Analyse des erweiterten Bezugsrahmenmodells

Modell	$\chi^2$	df	SF	CFI	RMSEA	SRMR
<i>Ohne Kontrolle der Schulform</i>						
Standardmodell	8.343	175	1,078	.97	.04	.06
Standardmodell-direkt	6.039	169	1,080	.98	.03	.03
Nested-Faktormodell	7.589	170	1,078	.98	.04	.05
Nested-Faktormodell-direkt	5.244	164	1,080	.98	.03	.03
<i>Mit Kontrolle der Schulform</i>						
Standardmodell	6.951	175	1,074	.97	.04	.04
Standardmodell-direkt	5.282	169	1,074	.98	.03	.03
Nested-Faktormodell	6.736	170	1,073	.97	.04	.04
Nested-Faktormodell-direkt	5.047	164	1,072	.98	.03	.03

SF = Skalierungsfaktor.

*Latente Korrelationen.* Bevor auf die Regressionskoeffizienten eingegangen wird, sollen zunächst die latenten Korrelationen betrachtet werden. In Tabelle 35 sind die latenten Korrelationen für das Standardmodell und in Tabelle 36 die Korrelationen für das Nested-Faktormodell eingetragen.

Mit Blick auf die latenten Korrelationen sind mehrere Aspekte interessant:

- Auch bei Kontrolle der Schulformzugehörigkeit blieb in beiden Modellen das Muster der Interkorrelationen insgesamt gesehen weitestgehend unverändert. Die Veränderungen der Korrelationen waren unter anderem eine Folge der Varianzrestriktion, die mit einer Kontrolle der Schulformzugehörigkeit einherging (siehe Anhang).
- Die fluide Fähigkeit wie auch die allgemeine kognitive Fähigkeit korrelierten jeweils positiv und jeweils in nahezu gleicher Höhe mit dem Interesse am Lesen und dem mathematischen Interesse. Die Stärke der Korrelation war als gering (in Bezug auf das mathematische Interesse) bis mittel (in Bezug auf das Interesse am Lesen) zu beurteilen.
- Die mathematikspezifische und die spezifische verbale Fähigkeit korrelierten jeweils positiv (in mittlerer Höhe) mit dem korrespondierenden fachspezifischen Interesse und jeweils negativ (in mittlerer Höhe) mit dem Interesse der jeweils anderen Domäne. Mit Blick auf dieses letzte Ergebnis zeigte sich – wie auch bei den Selbstkonzepten – ein deutlicher Unterschied zu den korrespondierenden Fähigkeiten im Standardmodell. Hier fanden sich geringfügig positive (in Bezug auf die generelle mathematische Fähigkeit) oder geringfügig negative (in Bezug auf die generelle verbale Fähigkeit) Korrelationen mit den Interessen der jeweils anderen Inhaltsdomäne.



Tabelle 35: Standardmodell: Latente Korrelationen der kognitiven Fähigkeiten mit den fachspezifischen Selbstkonzepten und Interessen

	M	Gf	V	INT-L	INT-M	SK-V	SK-M
M		.64	.79	.07	.35	-.03	.46
Gf	.89		.70	.31	-.02	.23	.09
V	.80	.83		.21	.19	.07	.30
INT-L	.19	.29	.36		-.04	.33	-.04
INT-M	.22	.10	-.05	-.05		-.14	.87
SK-V	.07	.13	.25	.35	-.15		-.13
SK-M	.34	.22	.06	-.04	.87	-.14	

Der Modell-Fit für das Modell zur Berechnung der Korrelationen war identisch mit dem Standardmodell-direkt (siehe Tab. 34). Unterhalb der Diagonale sind die Korrelationen ohne Kontrolle der Schulformzugehörigkeit eingetragen. Über der Diagonale stehen die latenten Korrelationen bei Kontrolle der Schulformzugehörigkeit. Der Modell-Fit war identisch zu den Werten, die in Tabelle 34 für das Standardmodell-direkt berichtet wurden. M = generelle mathematische Fähigkeit, Gf = fluide Fähigkeit, V = generelle verbale Fähigkeit, INT-L = Interesse am Lesen, INT-M = mathematisches Interesse, SK-V = verbales Selbstkonzept, SK-M = mathematisches Selbstkonzept.

Tabelle 36: Nested-Faktormodell: Latente Korrelationen der kognitiven Fähigkeiten mit den fachspezifischen Selbstkonzepten

	M'	g	V'	INT-L	INT-M	SK-V	SK-M
M'		.00 <sup>a</sup>	.00 <sup>a</sup>	-.22	.37	-.17	.39
g	.00 <sup>a</sup>		.00 <sup>a</sup>	.22	-.23	.27	-.20
V'	.00 <sup>a</sup>	.00 <sup>a</sup>		.22	.18	.07	.30
INT-L	-.22	.31	.20		-.04	.33	-.05
INT-M	.34	.09	-.26	-.05		-.14	.87
SK-V	-.17	.15	.26	.35	-.15		-.14
SK-M	.37	.21	-.24	-.05	.87	-.14	

Der Modell-Fit für das Modell zur Berechnung der Korrelationen war identisch mit dem Nested-Faktormodell-direkt (siehe Tab. 34). Unterhalb der Diagonale sind die Korrelationen ohne Kontrolle der Schulformzugehörigkeit eingetragen. Über der Diagonale stehen die latenten Korrelationen bei Kontrolle der Schulformzugehörigkeit. M' = mathematikspezifische Fähigkeit, g = allgemeine kognitive Fähigkeit, V' = spezifische verbale Fähigkeit, INT-L = Interesse am Lesen, INT-M = mathematisches Interesse, SK-V = verbales Selbstkonzept, SK-M = mathematisches Selbstkonzept.

<sup>a</sup> Diese Korrelationen wurden auf Null fixiert.

*Analyse der erweiterten Bezugsrahmenmodelle.* Die Modellparameter zur Auswertung des erweiterten Bezugsrahmenmodells in Form des Standardmodells sind in Tabelle 37 eingetragen. Das Befundmuster zur Vorhersage der fachspezifischen Selbstkonzepte war äquivalent zur Analyse des Bezugsrahmenmodells in Abschnitt 9.3.3. Weiterhin hatten die fachspezifischen Selbstkonzepte jeweils den erwarteten positiven Effekt auf das jeweilige korrespondierende Interesse (Tab. 37). Hier ist insbesondere das sehr hohe standardisierte Regressionsgewicht des mathematischen Selbstkonzepts hervorzuheben ( $\beta = .86$ ). Wider Erwarten hatte jedoch das mathematische Selbstkonzept nahezu keinen Effekt auf das Interesse am Lesen. Auch der negative Effekt des verbalen Selbstkonzepts auf das mathematische Interesse war nahezu zu vernachlässigen.

Tabelle 37: Standardmodell zur Prüfung des erweiterten Bezugsrahmenmodells: Standardisierte Regressionsgewichte und erklärte Varianz ( $R^2$ )

UV	Standardmodell				Standardmodell-direkt			
	Abhängige Variable				Abhängige Variable			
	SK-M	SK-V	INT-M	INT-L	SK-M	SK-V	INT-M	INT-L
SK-M			.86* (.87)	.01 (.00)			.86* (.88)	.03* (.04)
SK-V			-.04* (-.03)	.37* (.35)			.00 (.00)	.25* (.25)
V	-.56* (-.35)	.57* (.43)			-.54* (-.34)	.54* (.41)	-.13* (-.09)	.35* (.28)
M	.82* (.69)	-.42* (-.33)			.82* (.68)	-.39* (-.31)	.07* (.04)	-.37* (-.30)
Gf	-.06 (-.02)	.05 (.04)			-.05 (-.01)	.03 (.02)	-.05* (-.04)	.29* (.22)
$R^2$	.23 (.28)	.13 (.12)	.76 (.76)	.13 (.12)	.23 (.28)	.12 (.11)	.77 (.77)	.23 (.19)

\*  $p < .05$  (einseitiger Test).

In Klammern stehen die Modellparameter bei Kontrolle der Schulformzugehörigkeit. UV = unabhängige Variable, SK-M = mathematisches Selbstkonzept, SK-V = verbales Selbstkonzept, INT-M = mathematisches Interesse, INT-L = Interesse am Lesen, V = generelle verbale Fähigkeit, M = generelle mathematische Fähigkeit, Gf = fluide Fähigkeit. Die Korrelation zwischen den Residuen von SK-M und SK-V betrug im Standardmodell (Standardmodell-direkt)  $-.04$  ( $-.05$ ). Die Korrelation zwischen den Residuen von INT-M und INT-L betrug im Standardmodell (Standardmodell-direkt)  $.00$  ( $.03$ ).

Im Gegensatz zu den Selbstkonzepten hatte aber im Standardmodell-direkt die generelle mathematische Fähigkeit einen negativen Effekt in mittlerer Höhe auf das Interesse am Lesen und einen statistisch signifikanten, aber praktisch kaum bedeutsamen positiven Effekt auf das mathematische Interesse. Im Vergleich hierzu war der negative Effekt der generellen verbalen Fähigkeit auf das Interesse an Mathematik deutlich schwächer, hingegen der positive Effekt auf das Interesse am Lesen deutlich stärker. Die fluide Fähigkeit hatte nur auf das Interesse am Lesen einen praktisch bedeutsamen positiven Effekt.

Insgesamt gesehen konnte durch Hinzunahme der kognitiven Fähigkeiten auch die Varianzaufklärung beim Interesse am Lesen deutlich verbessert werden: Das  $R^2$  verbesserte sich von  $.13$  im Standardmodell auf  $.23$  im Standardmodell-direkt. Weiterhin führte die Aufhebung der Restriktion, die besagt, dass die kognitiven Fähigkeiten im Standardmodell die Interessen nicht direkt beeinflussen, zu einer signifikanten Verbesserung des Modell-Fits ( $\Delta\chi^2 = 2.419$ ,  $df = 6$ ,  $p < .05$ ).

berechnet mit den Daten aus Tabelle 34 und der Korrekturformel in Muthén & Muthén, 1998–2004a, S. 22).

Auch im Nested-Faktormodell (Tab. 38) waren die Befunde zur Vorhersage der fachspezifischen Selbstkonzepte wiederum äquivalent zur Analyse des Bezugsrahmenmodells in Abschnitt 9.3.3. Hinsichtlich der Vorhersage der fachspezifischen Interessen hatten die korrespondierenden Selbstkonzepte jeweils den erwarteten positiven Effekt. Wie auch im Standardmodell war hierbei der Effekt des mathematischen Selbstkonzepts besonders stark. Der erwartete negative Effekt auf das Interesse des anderen Fachs blieb aus.

Tabelle 38: Nested-Faktormodell zur Prüfung des erweiterten Bezugsrahmenmodells: Standardisierte Regressionsgewichte und erklärte Varianz ( $R^2$ )

UV	Nested-Faktormodell				Nested-Faktormodell-direkt			
	Abhängige Variable				Abhängige Variable			
	SK-M	SK-V	INT-M	INT-L	SK-M	SK-V	INT-M	INT-L
SK-M			.86* (.87)	.01 (.00)			.87* (.88)	.04* (.04)
SK-V			-.04* (-.03)	.37* (.35)			.00 (.00)	.25* (.25)
V'	-.25* (-.21)	.27* (.28)			-.24* (-.20)	.26* (.27)	-.06* (-.05)	.15* (.16)
M'	.37* (.40)	-.19* (-.18)			.37* (.39)	-.17* (-.17)	.03* (.03)	-.19* (-.19)
g	.19* (.29)	.17* (.09)			.21* (.30)	.15* (.07)	-.09* (-.08)	.26* (.19)
$R^2$	.23 (.28)	.13 (.12)	.76 (.76)	.13 (.12)	.23 (.28)	.12 (.11)	.77 (.77)	.24 (.20)

\*  $p < .05$  (einseitiger Test).

In Klammern stehen die Modellparameter bei Kontrolle der Schulformzugehörigkeit. UV = Unabhängige Variable, SK-M = mathematisches Selbstkonzept, SK-V = verbales Selbstkonzept, INT-M = mathematisches Interesse, INT-L = Interesse am Lesen, V' = spezifische verbale Fähigkeit, M' = mathematikspezifische Fähigkeit, g = allgemeine kognitive Fähigkeit. Die Korrelation zwischen den Residuen von SK-M und SK-V betrug im Nested-Faktormodell (Nested-Faktormodell-direkt)  $-.04 (-.04)$ . Die Korrelation zwischen den Residuen von INT-M und INT-L betrug im Nested-Faktormodell (Nested-Faktormodell-direkt)  $.00 (.03)$ .

Über die Selbstkonzepte hinaus wurde im Nested-Faktormodell-direkt das Interesse am Lesen negativ durch die mathematikspezifische Fähigkeit und jeweils positiv durch die spezifische verbale Fähigkeit und die allgemeine kognitive Fähigkeit beeinflusst. Im Gegensatz hierzu waren die Effekte der kognitiven Fähigkeiten auf das mathematische Interesse trotz ihrer statistischen Signifikanz praktisch kaum bedeutsam. Dies zeigte sich unter anderem auch daran, dass sich die aufgeklärte Varianz beim Interesse am Lesen von  $.13$  im Nested-Faktormodell auf  $.24$  im Nested-Faktormodell-direkt verbesserte, wohingegen die aufgeklärte Varianz beim mathematischen Interesse trotz der Hinzunahme der kognitiven Fähigkeiten als Prädiktoren nahezu unverändert blieb.

Insgesamt gesehen führte die Aufhebung der Restriktion, die besagt, dass die kognitiven Fähigkeiten im Standardmodell die Interessen nicht direkt beeinflussen, zu einer signifikanten Verbesserung des Modell-Fits ( $\Delta\chi^2 = 2.460$ ,  $df = 6$ ,  $p < .05$  berechnet mit den Daten aus Tabelle 34 und der Korrekturformel in Muthén & Muthén, 1998–2004a, S. 22).

Ein analoges Befundmuster resultierte für alle vier analysierten Modelle auch bei Kontrolle der Schulform. Die Unterschiede zwischen den standardisierten Modellparametern bei den Modellen mit und ohne Kontrolle der Schulformzugehörigkeit waren wohl in erster Linie auf die Varianzrestriktionen und den damit verbundenen Implikationen bei der Berechnung standardisierter Parameter zurückzuführen.

#### 9.4.4 Diskussion

Die zentralen Befunde der vorangegangenen Analysen können wie folgt zusammengefasst werden:

- Die fachspezifischen Selbstkonzepte hatten unabhängig von der Konzeptualisierung der kognitiven Fähigkeiten mathematischer Schülerleistung jeweils einen positiven Effekt auf das korrespondierende Interesse. Hier ist insbesondere der sehr starke Effekt des mathematischen Selbstkonzepts hervorzuheben. Erwartungswidrig hatten die Selbstkonzepte in keinem der analysierten Modelle einen praktisch bedeutsamen negativen Effekt auf das Interesse der jeweils anderen Inhaltsdomäne.
- Die kognitiven Fähigkeiten in den „direkt“-Varianten des Nested-Faktormodells und des Standardmodells konnten substantiell die Vorhersage beim Interesse am Lesen verbessern. Diese Verbesserung ging auf die positiven Effekte der allgemeinen kognitiven Fähigkeit (bzw. der fluiden Fähigkeit) und der spezifischen verbalen Fähigkeit (bzw. der generellen verbalen Fähigkeit) sowie den negativen Effekt der mathematikspezifischen Fähigkeit (bzw. der generellen mathematischen Fähigkeit) zurück.
- Die positiven und die negativen Effekte der spezifischen Fähigkeiten ( $M'$  und  $V'$ ) spiegelten sich auch in den bivariaten latenten Korrelationen mit den fachspezifischen Interessen wider: So korrelierte beispielsweise die mathematikspezifische Fähigkeit in geringfügiger bis mittlerer Höhe negativ mit dem Interesse am Lesen. Hingegen resultierte für die generelle mathematische Fähigkeit eine positive Korrelation mit dem Interesse am Lesen, die als gering bis mittel beurteilt werden konnte.

Zur Ableitung der Zusammenhänge wurde argumentiert, dass das fachspezifische Kompetenzerleben ein bedeutsamer Prädiktor für die Entwicklung der fachspezifischen Interessen ist. Zur Prüfung dieser Annahmen wurde unterstellt, dass die fachspezifischen Selbstkonzepte Indikatoren des Kompetenzerlebens sind und somit den fachspezifischen Interessen kausal vorgeschaltet sind. Höhere fachspezifische Selbstkonzepte gingen auch tatsächlich mit einem höheren korrespondierenden fachspezifischen Interesse einher. Im Vergleich zum verbalen Selbstkonzept und dem Interesse am Lesen war der Zusammenhang zwischen dem mathematischen Selbstkonzept und dem mathematischen Interesse sehr hoch. Als Erklärung bietet sich hierfür an, dass sich beide explizit auf die Mathematik(-leistung) bezogen. Hingegen thematisierte das verbale Selbstkonzept das Selbstkonzept in Deutsch und nicht das Selbstkonzept in Lesen. Letzteres mag auch der Grund dafür gewesen sein, dass die kognitiven Fähigkeiten über den sehr starken positiven Effekt des mathematischen Selbstkonzepts hinaus keinen substantiellen Effekt mehr auf das mathematische Interesse, wohl aber auf das Interesse am Lesen hatten.

In diesem Zusammenhang ist bemerkenswert, dass erwartungswidrig die negativen Effekte der Selbstkonzepte auf das Interesse des jeweils anderen Fachs nicht vorzufinden waren. Das heißt

zum Beispiel, dass für zwei Schüler mit gleichem verbalem Selbstkonzept es nahezu keinen Unterschied für das Interesse am Lesen machte, ob das mathematische Selbstkonzept hoch oder niedrig ausgeprägt war. Die Annahme im erweiterten Bezugsrahmenmodell, dass die fachspezifischen Selbstkonzepte einen negativen Effekt auf das Interesse der jeweils anderen Domäne haben, wurde somit von den Ergebnissen nicht überzeugend gestützt. Folglich wurde die negative bivariate Korrelation zwischen der mathematikspezifischen Fähigkeit und dem Interesse am Lesen nur teilweise durch den negativen indirekten Effekt über das verbale Selbstkonzept vermittelt. Darüber hinaus bestand aber immer noch ein direkter negativer Effekt der mathematikspezifischen Fähigkeit bzw. der generellen mathematischen Fähigkeit auf das Interesse am Lesen. Die fluide Fähigkeit und die allgemeine kognitive Fähigkeit hatten dabei jeweils positive Effekte. Eine weitergehende Interpretation dieses Befunds wird im Rahmen der Gesamtdiskussion gegeben.

Resümierend kann an dieser Stelle festgehalten werden, dass die Befunde der konvergenten und diskriminanten Validität zum fachspezifischen Interesse erneut deutlich machten, dass bei getrennter Betrachtung der kognitiven Fähigkeiten mathematischer Schülerleistung ein differenziertes Bild über ein zentrales Konstrukt der pädagogischen Psychologie – dem fachspezifischen Interesse – gezeichnet werden konnte: Insbesondere beim Vergleich der bivariaten Korrelationen zwischen dem Interesse am Lesen und der generellen mathematischen Fähigkeit sowie der mathematikspezifischen Fähigkeit wurde dies offensichtlich. Bei einer Konzeptualisierung der kognitiven Fähigkeiten mathematischer Schülerleistung in Form des Standardmodells wurden diese Beziehungen erst durch die regressionsanalytische Kontrolle der beiden anderen kognitiven Fähigkeiten offensichtlich (siehe auch Abschnitt 9.3 und 10.3.1).

## 9.5 Schulnoten

### 9.5.1 Empirische Befundlage und Hypothesen

Die beiden vorangegangenen Abschnitten befassten sich mit dem Zusammenhang zwischen den kognitiven Fähigkeiten mathematischer Schülerleistung und der Lernmotivation. Dieser Abschnitt beschäftigt sich aus zwei Gründen mit Schulnoten:

- (a) Für die Entwicklung der fachspezifischen Lernmotivation sind Schulnoten zentral, da sie den Schülern Rückmeldung über ihre fachspezifischen Fähigkeiten geben. Dies wird belegt durch empirische Studien zum Bezugsrahmenmodell (z.B. Lüdtke u.a., 2002; für eine Zusammenfassung siehe Möller & Köller, 2004). Die Rückmeldung der eigenen Leistungen trägt auch zum Kompetenzerleben der Schüler bei. Aufgrund des anzunehmenden Zusammenhangs zwischen Kompetenzerleben und Interesse sind die Schulnoten daher auch für die Interessensentwicklung relevant.
- (b) Unabhängig von ihrer Rolle bei der Genese der Lernmotivation stellen Schulnoten eines der „beliebtesten“ Außenkriterien zur Validierung von standardisierten Leistungs- oder Intelligenztests dar, denn „the field of intelligence testing was born from the need to develop a test that would predict children’s school success“ (Mayer, 2000, S. 519).

Zusammenfassend lässt sich also festhalten, dass Schulnoten für die Entwicklung der Lernmotivation sehr bedeutsam sind und ein zentrales Außenkriterium zur Validierung von Leistungstests darstellen. Zur konvergenten Validierung der kognitiven Fähigkeiten mathematischer Schülerleistung werden in diesem Abschnitt die korrelativen Beziehungen mit der Mathematiknote (M-

Note) der Schüler analysiert. Die diskriminante Validität wird anhand der Deutschnote (D-Note) untersucht. Hinsichtlich der Mathematiknote werden die folgenden Hypothesen geprüft:

*Hypothesen 5a und 5b.* Entsprechend den Annahmen der Modelle schulischen Lernens (vgl. Abschnitt 5.1) ist die allgemeine kognitive Fähigkeit ein bedeutsamer Prädiktor der fachspezifischen Lernleistung. Daher kann erwartet werden, dass die allgemeine kognitive Fähigkeit mit besseren Noten in Mathematik und in Deutsch einhergeht. Noten werden in Deutschland in der Regel mit „1 = sehr gut“ bis „6 = ungenügend“ codiert. Das bedeutet, dass höhere Werte in der allgemeinen kognitiven Fähigkeit mit niedrigeren Fachnoten assoziiert und somit substantziell negativ korreliert sind (Hypothese 5a:  $r_{g, M\text{-}Note} < 0$ ; Hypothese 5b:  $r_{g, D\text{-}Note} < 0$ ).

*Hypothesen 5c und 5d.* Ebenso wie die allgemeine kognitive Fähigkeit sollte die fluide Fähigkeit mit besseren Noten in Mathematik und in Deutsch einhergehen (Hypothese 5c:  $r_{Gf, M\text{-}Note} < 0$ ; Hypothese 5d:  $r_{Gf, D\text{-}Note} < 0$ ).

Die Hypothese zur negativen Korrelation zwischen Schulnoten und allgemeiner Intelligenz bzw. fluider Fähigkeit wird durch zahlreiche Einzelstudien (z.B. Gustafsson & Balke, 1993; Süß, 2001) empirisch gestützt. Jensen spricht davon, dass „If there is any unquestioned fact in applied psychometrics, it is that IQ tests [damit meint Jensen Tests zur Messung kognitiver Fähigkeiten; M.B.] have a high degree of predictive validity for many educational criteria, such as (...) school and college grades.“ (Jensen, 1998, S. 277)

Nach Tent (2001, S. 808) liegt die Korrelation zwischen der Note in Mathematik und der allgemeinen kognitiven Fähigkeit bei ungefähr  $-0.50$ . Für andere Fächer liegt die Korrelation mit der allgemeinen kognitiven Fähigkeit etwas unter diesem Wert.

*Hypothesen 5e und 5f.* Analog zur allgemeinen kognitiven Fähigkeit (bzw. fluiden Fähigkeit) sollten auch die generelle mathematische Fähigkeit und die mathematikspezifische Fähigkeit Prädiktoren des Schulerfolgs in Mathematik sein (z.B. Gruehn, 2000; Köller & Baumert, 2001; Kunter, 2005) und somit mit besseren Mathematiknoten einhergehen (Hypothese 5e:  $r_{M, M\text{-}Note} < 0$ ; Hypothese 5f:  $r_{M', M\text{-}Note} < 0$ ).

In der Forschungsliteratur findet man häufig Korrelationen zwischen Schultests und Fachnoten, die zwischen  $-0.60$  und  $-0.70$  liegen (Tent, 2001, S. 808).

Ob die generelle mathematische Fähigkeit oder die mathematikspezifische Fähigkeit stärker mit der Mathematiknote assoziiert ist, kann a priori nicht bestimmt werden. Diese Frage wird explorativ untersucht.

Wie korrelieren die mathematischen Fähigkeiten mit der Deutschnote? Zunächst wird der Zusammenhang zwischen der mathematikspezifischen Fähigkeit und der Deutschnote betrachtet:

*Hypothese 5g.* Es wird erwartet, dass die mathematikspezifische Fähigkeit nicht mit der Deutschnote korreliert ( $r_{M', D\text{-}Note} = 0$ ). Es ist nicht plausibel, dass die mathematikspezifische Fähigkeit, die in dieser Arbeit als spezifische Fähigkeit zum Lösen mathematischer Probleme interpretiert wird (vgl. Abschnitt 7.3), ein bedeutsamer Prädiktor für die Leseleistung ist. Da die mathematikspezifische Fähigkeit nicht leistungsrelevant ist, sollte sie auch nicht mit der Deutschnote korrelieren.

*Hypothese 5b.* Laut Hypothese 5b ist die allgemeine kognitive Fähigkeit negativ mit der Deutschnote korreliert. Allerdings korreliert entsprechend der obigen Annahme die mathematikspezifische Fähigkeit nicht mit der Deutschnote. Somit sollte zusammengenommen (angesichts der „Mischung“ von  $g$  und  $M'$  in der generellen mathematischen Fähigkeit) insgesamt eine negative „Netto“-Korrelation zwischen der generellen mathematischen Fähigkeit und der Deutschnote resultieren ( $r_{M, D-Note} < 0$ ).

*Hypothese 5i.* Folgt man den Annahmen von Hypothese 5g, dann leuchtet ein, dass die Korrelation zwischen der generellen mathematischen Fähigkeit und der Deutschnote (für die ein negativer Wert erwartet wird) kleiner sein sollte als die Korrelation zwischen der mathematikspezifischen Fähigkeit und der Deutschnote ( $r_{M, D-Note} < r_{M', D-Note}$ ). Denn von Letzterer wird angenommen, dass sie Null ist.

Zum Zusammenhang zwischen der Deutschnote und mathematischen Fähigkeiten gibt es meines Wissens keine systematischen Zusammenfassungen der korrespondierenden Korrelationen. Lütke und Kollegen berichteten eine Korrelation von  $-0.16$  zwischen der Deutschnote und der generellen mathematischen Fähigkeit (für eine Stichprobe von 15-Jährigen, die an PISA 2000 teilgenommen hatten).

## 9.5.2 Methode

*Manifeste Variablen.* Die Schulnoten wurden durch den Schülerfragebogen erfasst, der bei PISA 2000 eingesetzt wurde (Kunter u.a., 2002). Hierzu beantworteten die Schüler die Frage „Welche Zensuren hattest du im letzten Zeugnis in folgenden Fächern?“ unter anderem für die Fächer Mathematik und Deutsch mithilfe einer Notenskala (1 = sehr gut, 2 = gut, 3 = befriedigend, 4 = ausreichend, 5 = mangelhaft, 6 = ungenügend). Negative Korrelationen der kognitiven Fähigkeiten mit den Schulnoten bedeuteten also, dass höhere Werte in der jeweiligen kognitiven Fähigkeit mit besseren Schulnoten einhergingen.

Um für schulspezifische Unterschiede in der Handhabung der Notengebung (siehe hierzu z.B. Baumert, Trautwein & Artelt, 2003; Lütke u.a., 2002) zu kontrollieren, wurden für die Analysen in dieser Arbeit die Mathematiknote und die Deutschnote am jeweiligen Schulmittelwert zentriert. Insgesamt fehlten von 1.394 Schülern die Angaben zur Deutschnote und von 1.397 Schülern die Angaben zur Mathematiknote.

*Analyseverfahren.* Die statistische Analysestrategie war identisch mit der in Abschnitt 9.1.2 beschriebenen Prozedur. In getrennten Analysen wurden die korrelativen Beziehungen der kognitiven Fähigkeiten zur Mathematiknote und zur Deutschnote untersucht. Für die Noten in beiden Fächern wurden wie bei der Analyse der Geschlechterunterschiede latente Notenvariablen spezifiziert (vgl. Abb. 26a und 26b). Zur Identifikation wurde jeweils die latente Varianz auf 1,0 und die Residualvarianz der manifesten Notenvariablen auf 0 fixiert. Die Faktorladungen der manifesten Notenvariablen wurden frei geschätzt.

*Kontrolle der Schulformzugehörigkeit.* Die korrelativen Beziehungen zwischen den Schulnoten und den kognitiven Fähigkeiten wurden auch bei Kontrolle der Schulformzugehörigkeit analysiert, um für mögliche Schulformeffekte zu kontrollieren (siehe Abschnitt 9.1.2 für die Beschreibung der Methodik).

### 9.5.3 Ergebnisse

In Tabelle 39 sind die Korrelationen zwischen den Schulnoten und den kognitiven Fähigkeiten eingetragen (siehe auch Abb. 30a und 30b in Abschnitt 9.6 für eine grafische Übersicht der Validitätskoeffizienten). Zunächst werden die Beziehungen zur allgemeinen kognitiven Fähigkeit und zur fluiden Fähigkeit behandelt.

*Zusammenhang der Noten mit g sowie Gf.* Je stärker die allgemeine kognitive Fähigkeit sowie die fluide Fähigkeit eines Schülers ausgeprägt waren, desto besser war seine Note in Mathematik und in Deutsch. Die Stärke der Korrelationen war als geringfügig bis mittel zu beurteilen. Wie zu erwarten, waren diese Korrelationen auch signifikant von Null verschieden (Tab. 39). Die Hypothesen 5a, 5b, 5c und 5d wurden somit deskriptiv und inferenzstatistisch von den empirischen Befunden gestützt.

Tabelle 39: Korrelationen ( $r$ ) zwischen kognitiven Fähigkeiten (KF) und Schulnoten sowie korrespondierende  $\chi^2$ -Differenzentests ( $\Delta\chi^2$ ) zur Prüfung der Hypothesen (Hpt)

Hpt	KF	$r$	$r_{\text{Schulform}}$	Test	$\chi^2$ (df = 61)	SF	$\Delta\chi^2$
<i>Mathematiknote</i>							
5a	g	-.23	-.30	$r_{g, M\text{-Note}} = 0$	1.273	1,558	683,3*
5c	Gf	-.23	-.30	$r_{Gf, M\text{-Note}} = 0$	1.755	1,559	648,8*
5e	M	-.31	-.39	$r_{M, M\text{-Note}} = 0$	2.421	1,561	1.167,5*
5f	M'	-.24	-.25	$r_{M', M\text{-Note}} = 0$	841	1,550	375,9*
	M & M'			$r_{M, M\text{-Note}} = r_{M', M\text{-Note}}$	590	1,547	35,8*
<i>Deutschnote</i>							
5b	g	-.16	-.19	$r_{g, D\text{-Note}} = 0$	836	1,555	397,1*
5d	Gf	-.15	-.18	$r_{Gf, D\text{-Note}} = 0$	1.283	1,556	345,9*
5g	M'	.10	.06	$r_{M', D\text{-Note}} = 0$	566	1,551	75,5*
5h	M	-.10	-.13	$r_{M, D\text{-Note}} = 0$	1.144	1,552	219,2*
5i	M & M'			$r_{M, D\text{-Note}} = r_{M', D\text{-Note}}$	709	1,551	294,6*

\*  $p < .05$ .

$r_{\text{Schulform}}$  = Korrelation bei Kontrolle der Schulformzugehörigkeit, SF = Skalierungsfaktor. Zur Berechnung der  $\chi^2$ -Differenz ( $\Delta\chi^2$ ) wurde die Korrekturformel aus dem technischen Anhang von Mplus verwendet (Muthén & Muthén, 1998–2004a, S. 22). Alle  $\chi^2$ -Differenzen werden mit df = 1 getestet. Der  $\chi^2$ -Goodness-of-Fit-Testwert des unrestringierten Standardmodells, mit dem die jeweilige Parameterrestriktion hinsichtlich der Mathematiknote verglichen wurde, betrug  $\chi^2 = 1.009$  mit df = 60 und SF = 1.559 (Nested-Faktormodell:  $\chi^2 = 535$ , df = 60, SF = 1.559). Der korrespondierende Wert im Standardmodell für die Zusammenhangsanalysen mit der Deutschnote war  $\chi^2 = 980$  mit df = 60 und SF = 1.563 (Nested-Faktormodell:  $\chi^2 = 512$ , df = 60, SF = 1.563). Fähigkeiten im Standardmodell: Gf = fluide Fähigkeit, M = generelle mathematische Fähigkeit. Fähigkeiten im Nested-Faktormodell: g = allgemeine kognitive Fähigkeit, M' = mathematikspezifische Fähigkeit.

*Zusammenhang der Noten mit M sowie M'.* Höhere Leistungen bei den mathematischen Fähigkeiten gingen mit besseren Mathematiknoten einher. Die Korrelationen waren im Vergleich zu den Zusammenhängen der Noten mit der fluiden Fähigkeit bzw. der allgemeinen kognitiven Fähigkeit höher und konnten als tendenziell mittlerer Effekt betrachtet werden. Wie zu erwarten, waren



diese Korrelationen auch signifikant von Null verschieden. Die Hypothesen 5e und 5f wurden somit deskriptiv und inferenzstatistisch von den empirischen Befunden gestützt.

Der Zusammenhang zwischen der Mathematiknote und der generellen mathematischen Fähigkeit war etwas straffer als für die mathematikspezifische Fähigkeit. Dieser Unterschied des korrelativen Zusammenhangs war auch statistisch signifikant.

Wie korrelierten die mathematischen Fähigkeiten mit der Deutschnote? Bessere Leistungen bei der generellen mathematischen Fähigkeit gingen wie erwartet mit besseren Noten in Deutsch einher. Allerdings waren höhere Leistungen in der mathematikspezifischen Fähigkeit erwartungswidrig mit schlechteren Deutschnoten assoziiert. Beide Korrelationen waren statistisch signifikant von Null verschieden. Dieses Befundmuster stützte Hypothese 5h, falsifizierte aber Hypothese 5g.

Ein Vergleich der Korrelationskoeffizienten zeigte, dass sich die Zusammenhänge der mathematischen Fähigkeiten zur Deutschnote deutlich unterschieden. Die Differenz zwischen den beiden Korrelationen betrug .20. Diese Differenz war auch statistisch signifikant. Der deskriptive Befund wie auch die inferenzstatistische Analyse sprachen für Hypothese 5i.

Wie aus Tabelle 39 hervorgeht, resultierte ein analoges Befundmuster, wenn für die Schulformzugehörigkeit kontrolliert wurde. Abweichungen in der Höhe der Korrelationen sind (unter anderem) sicherlich auf die Varianzrestriktionen in den Maßen der kognitiven Fähigkeiten zu attribuieren.

#### 9.5.4 Diskussion

Die wichtigsten Ergebnisse zum Zusammenhang der kognitiven Fähigkeiten mit der Mathematiknote und der Deutschnote lassen sich wie folgt zusammenfassen:

- Höhere Leistungen bei der allgemeinen kognitiven Fähigkeit sowie bei der fluiden Fähigkeit waren hypothesenkonform mit besseren Mathematik- und Deutschnoten assoziiert (Hypothesen 5a, 5b, 5c und 5d). Der Zusammenhang zwischen den Noten und diesen beiden kognitiven Fähigkeiten war zudem gleich stark ausgeprägt.
- Höhere Leistungen bei den mathematischen Fähigkeiten gingen mit besseren Mathematiknoten einher (Hypothesen 5e und 5f). Jedoch zeigte sich ein differenzielles Befundmuster beim Vergleich des Zusammenhangs mit der Mathematiknote wie auch mit der Deutschnote (vgl. Hypothese 5i): Die Korrelation zwischen der generellen mathematischen Fähigkeit und der Mathematiknote war stärker als die korrespondierende Korrelation mit der mathematikspezifischen Fähigkeit. Während höhere Werte bei der generellen mathematischen Fähigkeit hypothesenkonform mit besseren Deutschnoten einhergingen (Hypothese 5h), war eine stärker ausgeprägte mathematikspezifische Fähigkeit entgegen der Erwartung mit schlechteren Deutschnoten assoziiert (Hypothese 5g).

Die Befunde wiesen klar darauf hin, dass sowohl die allgemeine kognitive Fähigkeit als auch die mathematikspezifische Fähigkeit „ökologische Validität“ im Unterricht besaßen: Höhere Leistungen in diesen beiden Fähigkeiten gingen mit besseren Mathematiknoten einher. Zwei Interpretationen des Nested-Faktormodells sind dabei interessant: Zum einen kann das Modell so interpre-

tiert werden, dass Schüler gleicher allgemeiner kognitiver Fähigkeit, aber mit höherer mathematikspezifischer Fähigkeit bessere Noten in Mathematik hatten. Zum anderen deutet das Befundmuster darauf hin, dass Schüler die gleiche Note in Mathematik auf unterschiedlichen Wegen erreichen konnten. Potenzielle Defizite in einer der beiden kognitiven Fähigkeiten wurden durch die jeweils andere Fähigkeit kompensiert: So war es zum Beispiel möglich, trotz niedriger Leistungen in der allgemeinen kognitiven Fähigkeit bei gleichzeitig hohen Leistungen in der mathematikspezifischen Fähigkeit dennoch eine gute Note in Mathematik zu erzielen.

Im Vergleich zu den Korrelationen zwischen kognitiven Fähigkeiten und Schulnoten, die in der Forschungsliteratur berichtet werden (Jensen, 1998; Tent, 2001), wurden in dieser Arbeit niedrigere Zusammenhänge gefunden. Dieses Ergebnis überraschte aber nicht so sehr, da bereits Lüdtke und Kollegen (Lüdtke u.a., 2002, Tab. 1) für 15-Jährige, die an der PISA-2000-Untersuchung teilgenommen hatten, Korrelationen in ähnlicher Höhe berichteten. Hinsichtlich der mathematischen Fähigkeiten war möglicherweise die Überlappung zwischen dem PISA-Mathematiktest und den Lehrplaninhalten für das erste Halbjahr der 9. Klasse, die der Notengebung zu Grunde lagen, nicht optimal. Denn die Zusammenstellung des PISA-Mathematiktests orientiert sich an einer Grundbildungskonzeption und nicht an einer Optimierung der Lehrplaninhalte für Neuntklässler. Mit Blick auf die allgemeine kognitive Fähigkeit ist ein Grund für die niedrigeren Korrelationen darin zu sehen, dass sich die in der Forschungsliteratur berichteten Korrelationen meist auf eine Gesamtnote beziehen und nicht auf Einzelnoten. Jedoch können sich Spezifika der jeweiligen Fächer durch die Aggregation von Einzelnoten zu einer Gesamtnote ausbalancieren (Little, Lindenberger & Nesselrode, 1999; Wittmann, 1988). So nimmt durch die Aggregation der Grad der „Symmetrie“ (Wittmann, 1988) zwischen der allgemeinen kognitiven Fähigkeit und der Gesamtnote zu, was letztlich mit einer Zunahme der Interkorrelation einhergeht.

Wie kann der Befund erklärt werden, dass die generelle mathematische Fähigkeit stärker mit der Mathematiknote korrelierte als die mathematikspezifische Fähigkeit? Eine mögliche Erklärung hierfür ist (vgl. Abschnitt 1.3.4), dass in die generelle mathematische Fähigkeit die allgemeine kognitive Fähigkeit und die mathematikspezifische Fähigkeit gemischt, aber *nicht* additiv eingehen. Es kann spekuliert werden, ob nicht ein Suppressoreffekt zwischen diesen beiden Fähigkeiten dazu führte, dass irrelevante Varianz in der generellen mathematischen Fähigkeit hinsichtlich der Korrelation mit der Mathematiknote unterdrückt wurde. Ein Indiz hierfür waren die standardisierten Regressionsgewichte ( $\beta$ ) in einer multiplen Regression zur Vorhersage der Mathematiknote mit der generellen mathematischen Fähigkeit und der fluiden Fähigkeit als Prädiktoren. Die generelle verbale Fähigkeit konnte dabei mit der Residualvarianz der Mathematiknote korrelieren. Während höhere Leistungen bei der generellen mathematischen Fähigkeit eine bessere Mathematiknote vorhersagten ( $\beta = -.46$ ), waren höhere Leistungen in der fluiden Fähigkeit (trotz positiver Korrelation mit der Mathematiknote) mit schlechteren Mathematiknoten assoziiert ( $\beta = .17$ ). Bei Äquivalenz der allgemeinen kognitiven Fähigkeit und der fluiden Fähigkeit stützt dieser Befund die Annahme einer Suppressoreffekt. Unabhängig davon sollte allerdings der relativ geringe Unterschied des korrelativen Zusammenhangs von .07 nicht überinterpretiert werden.

Interessanter als dieser mögliche Suppressoreffekt war aus meiner Sicht jedoch das zweite erwartungswidrige Ergebnis: Höhere Werte der mathematikspezifischen Fähigkeit waren mit schlechteren Deutschnoten assoziiert. Dieser Befund wird im Rahmen der Gesamtdiskussion weiter erörtert.

## 9.6 Zusammenfassende Diskussion zur Validität

Ziel der dritten Forschungsfrage war es, ein genaueres Bild der Zusammenhänge zwischen den kognitiven Fähigkeiten und soziodemografischen und motivationalen Schülermerkmalen sowie Schulnoten zu zeichnen. Einen Überblick über die zentralen Ergebnisse gibt Abbildung 30, in der die Validitätskoeffizienten für die allgemeine kognitive Fähigkeit und die fluide Fähigkeit (Abb. 30a) sowie für die generelle mathematische Fähigkeit und die mathematikspezifische Fähigkeit (Abb. 30b) dargestellt sind. Diese Validitätskoeffizienten wurden alle nach der in Abschnitt 9.1.2 beschriebenen Methode berechnet. Weiterhin sind im Anhang die Interkorrelationen aller untersuchten Schulervariablen (inkl. der Schulformzugehörigkeit) sowie einige zusammenfassende Regressionsanalysen dargestellt.

Vor dem Hintergrund dieser Ergebnisse werden nachfolgend zwei Themen diskutiert:

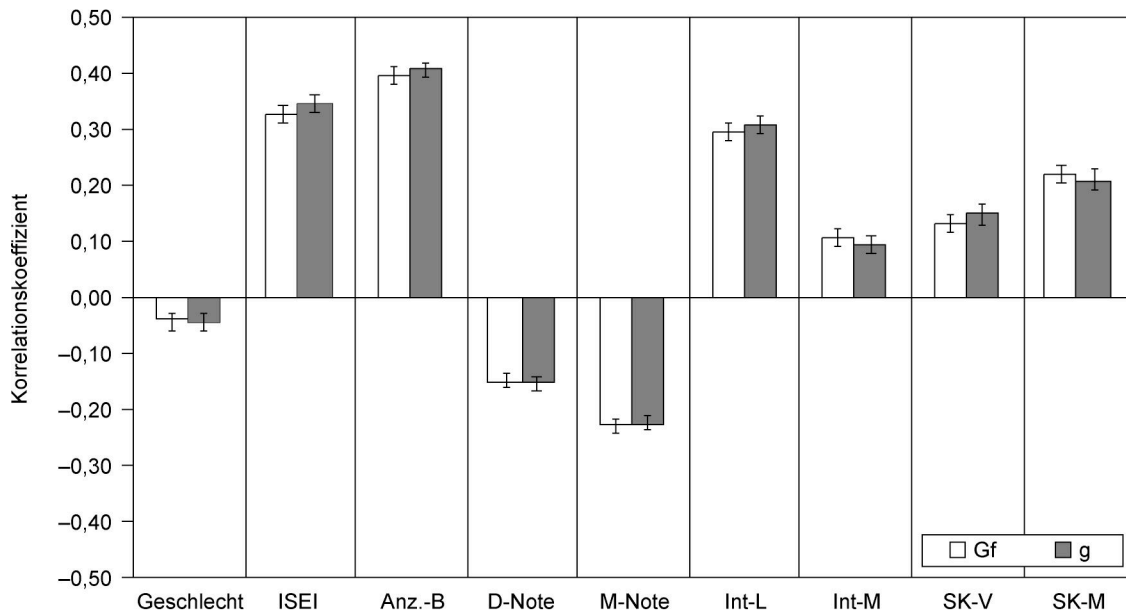
- Implikationen der Ergebnisse der dritten Forschungsfrage für Forschungsmethodik und Bewertung bisheriger Validitätsstudien von Maßen mathematischer Schülerleistung.
- Theoretische Integration kognitiver Fähigkeiten und motivationaler Schülermerkmale.

*Implikationen der Ergebnisse.* Aufschlussreich an Abbildung 30 ist der Vergleich der Korrelationen der Außenkriterien mit der fluiden Fähigkeit und der allgemeinen kognitiven Fähigkeit bzw. der generellen mathematischen Fähigkeit mit der mathematikspezifischen Fähigkeit. So waren die korrelativen Beziehungen zwischen den Außenkriterien und der fluiden Fähigkeit oder der allgemeinen kognitiven Fähigkeit nahezu identisch (Abb. 30a). Dies war zu erwarten, da mit Gustafsson (1984) angenommen wurde, dass diese beiden kognitiven Fähigkeiten äquivalent sind. Für die Analyse der Beziehungen zwischen der allgemeinen kognitiven Fähigkeit oder der fluiden Fähigkeit deuteten die Ergebnisse in dieser Arbeit also darauf hin, dass es keinen allzu großen Unterschied macht, ob man diese Fähigkeiten in Form eines Standardmodells oder eines Nested-Faktormodells konzeptualisiert.

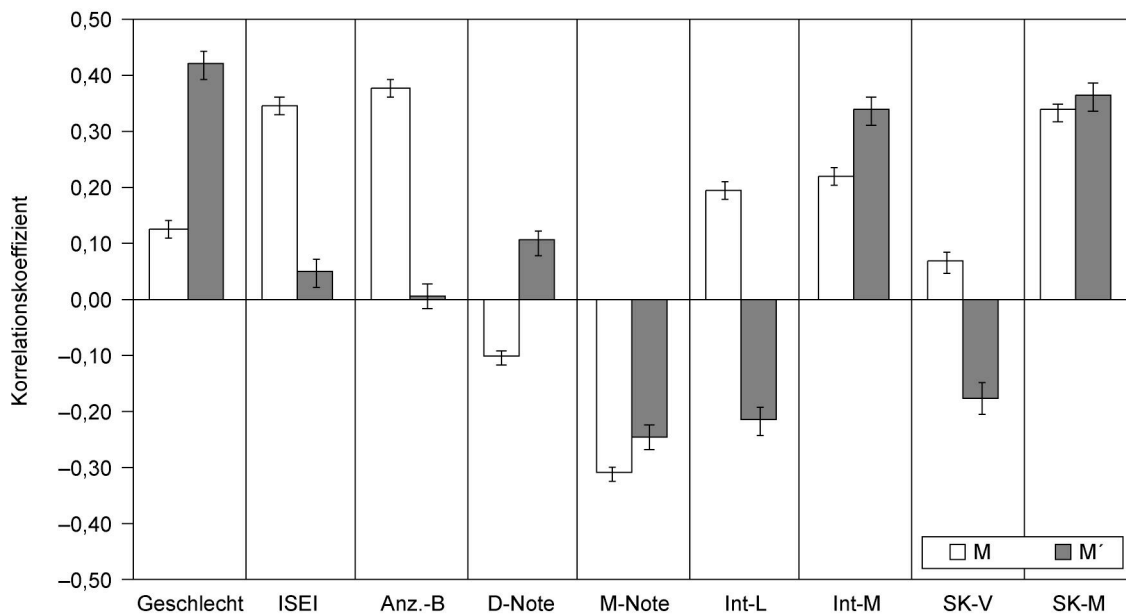
Im Gegensatz hierzu macht Abbildung 30b deutlich, dass es einen Unterschied machte, wie man die mathematischen Fähigkeiten konzeptualisierte. Es wurde argumentiert, dass in die generelle mathematische Fähigkeit die allgemeine kognitive Fähigkeit und die mathematikspezifische Fähigkeit gemischt eingingen. Immer dann, wenn diese beiden kognitiven Fähigkeiten differenziell valide hinsichtlich der untersuchten Schülermerkmale waren, waren einige Befunde bemerkenswert:

- Der Leistungsvorsprung zu Gunsten der Jungen war deutlich stärker bei der mathematikspezifischen Fähigkeit als bei der generellen mathematischen Fähigkeit ausgeprägt.
- Die mathematikspezifische Fähigkeit war nahezu unabhängig von Unterschieden des familiären Hintergrunds.
- Höhere Werte in der mathematikspezifischen Fähigkeit waren mit einem geringeren verbalen Selbstkonzept und einem geringeren Interesse am Lesen assoziiert.
- Je höher die mathematikspezifische Fähigkeit eines Schülers ausgeprägt war, desto schlechter war seine Note in Deutsch.

(a) Fluide Fähigkeit (Gf) und allgemeine kognitive Fähigkeit (g)



(b) Generelle mathematische Fähigkeit (M) und mathematikspezifische Fähigkeit (M')



ISEI = *International Socio-Economic Index of Occupational Status*, Anz.-B = Anzahl an Büchern, D-Note = Deutschnote, M-Note = Mathematiknote, Int-L = Interesse am Lesen, Int-M = mathematisches Interesse, SK-V = verbales Selbstkonzept, SK-M = mathematisches Selbstkonzept.

Abbildung 30: Korrelationen (mit 95%-Konfidenzintervall) zwischen kognitiven Fähigkeiten und Schülermerkmalen

Diese Befunde haben wie eingangs angekündigt wichtige Implikationen für die Bewertung der Validitätskoeffizienten bisheriger Forschung. Ist es nicht möglich, dass die berichteten Validitätskoeffizienten (in Abhängigkeit von der differenziellen Validität der allgemeinen kognitiven Fähigkeit und der mathematikspezifischen Fähigkeit) sowohl positiv als auch negativ „verschätzt“

worden sind? Dies wäre zumindest die Schlussfolgerung, die man ziehen müsste, wenn man die Konzeptualisierung der kognitiven Fähigkeiten mathematischer Schülerleistung in Form des Nested-Faktormodells favorisiert.

Es lohnt sich zu wiederholen, dass es hinsichtlich der mathematischen Fähigkeiten – im Gegensatz zur fluiden Fähigkeit und der allgemeinen kognitiven Fähigkeit – einen entscheidenden Unterschied machen kann, wie man interindividuelle Unterschiede in Maßen mathematischer Schülerleistung in Form von Strukturmodellen repräsentiert. Möchte man also einen differenzierteren Blick auf die Zusammenhänge zwischen verschiedenen Außenkriterien und Maßen mathematischer Schülerleistung werfen, ist es sehr empfehlenswert (zusätzlich zum Standardmodell) ein Nested-Faktormodell zu spezifizieren. Somit ist es möglich, die differenzielle Validität der allgemeinen kognitiven Fähigkeit und der mathematikspezifischen Fähigkeit zu analysieren. Dies impliziert jedoch, Maße fluider Fähigkeiten in das Studiendesign zu integrieren, denn erst dies ermöglicht eine eindeutige Interpretation des Generalfaktors als allgemeine kognitive Fähigkeit. Folgt man diesem Vorschlag, dann ermöglicht es auch eine profunde empirische und theoretisch gestützte Integration von kognitiven Fähigkeiten und motivationalen Schülermerkmalen. Deren potenzielles Zusammenwirken wird im nächsten Abschnitt beschrieben.

*Theoretische Integration kognitiver Fähigkeiten und motivationaler Schülermerkmale.* Von besonderem theoretischem Interesse sind die Ergebnisse zur mathematikspezifischen Fähigkeit. Wenn zwei Schüler die gleiche allgemeine kognitive Fähigkeit besaßen, sich aber in der mathematikspezifischen Fähigkeit unterschieden, hatte der Schüler mit der höheren mathematikspezifischen Fähigkeit in der Regel ein höheres mathematisches Selbstkonzept, war mathematisch interessierter und hatte bessere Noten in Mathematik. Das ist so zu erwarten. Allerdings hatte der Schüler mit der höheren mathematikspezifischen Fähigkeit auch ein niedrigeres verbales Selbstkonzept und war weniger interessiert am Lesen. Diese zunächst etwas überraschenden Befunde konnten durch das (erweiterte) Bezugsrahmenmodell und den internalen Vergleich zumindest teilweise erklärt werden. Jedoch erklärt dieses Modell nicht, weshalb für den Schüler mit der höheren mathematikspezifischen Fähigkeit schlechtere Noten in Deutsch zu erwarten waren. Dieser Befund kann durch das Zusammenspiel kognitiver Fähigkeiten und der Lernmotivation in Form von Selektions- und Investitionsprozessen erklärt werden.

Eine zentrale Annahme hierbei ist, dass Schüler weder über unbegrenzte kognitive noch unbegrenzte zeitliche Ressourcen verfügen. Diese Ressourcenknappheit kann das Lernverhalten von Schülern steuern. In Anlehnung an Ackermans (1996) *Intelligence-as-Process, Personality, Interests, and Intelligence-as-Knowledge*-Theorie (PPIK-Theorie) kann davon ausgegangen werden, dass die eigenen kognitiven Fähigkeiten ein bedeutsamer Prädiktor für den Erfolg und damit auch für das Kompetenzerleben in einem bestimmten Unterrichtsfach sind. Entsprechend der PPIK-Theorie (und natürlich auch den Erwartung-Wert-Modellen der Lernmotivation) ist die fächerspezifische Lernmotivation entscheidend dafür, welche Aufgaben und Probleme aus welchem Fachgebiet bearbeitet werden. Die erfolgreiche Bearbeitung einer Aufgabe steigert dann einerseits (vermittelt über Kompetenzerleben) die fachspezifische Lernmotivation, andererseits werden dadurch auch Fähigkeiten optimiert und neues Wissen erworben. So spricht Ackerman davon, dass „abilities and interests develop in tandem. (...) Thus subsequent to successful attempts at task performance, interest in the task domain may increase, along with the knowledge level for that task.“ (Ackerman, 1996, S. 244–245)

Unter der Annahme begrenzter zeitlicher und kognitiver Ressourcen kann man mit Ackerman davon ausgehen, dass Schüler abwägen müssen, in welche Unterrichtsfächer sie ihre Ressourcen

investieren (siehe hierzu auch Daniels, 2005). Entsprechend der Selbstbestimmungstheorie (z.B. Ryan, 1995) haben Menschen (neben dem Bedürfnis nach Autonomie und sozialer Eingebundenheit) das *essenzielle* Bedürfnis nach Kompetenzerleben. Daher sollten Schüler in die Fächer investieren, in denen sie die höchsten Aussichten auf Erfolg und Kompetenzerleben haben. Dies impliziert, dass die Investition in den Erwerb und die Optimierung der mathematikspezifischen Fähigkeit einerseits (unter anderem) mit einem gesteigerten mathematischen Selbstkonzept und mathematischem Interesse einhergeht. Andererseits geht für diese Schüler diese Art der Investition (wahrscheinlich mit Ausnahme hoch begabter Jugendlicher, vgl. Ackerman, 1996, S. 245) damit einher, dass sie aufgrund der begrenzten Ressourcen weniger Gelegenheiten besitzen, sich kompetent in anderen Fächern zu erleben. Das heißt, die motivationale Entwicklung in einem Unterrichtsfach geht potenziell zu Lasten der motivationalen Entwicklung in einem anderen Unterrichtsfach.

Dieser Mechanismus wird wahrscheinlich noch durch internale Vergleichsprozesse der eigenen Fähigkeiten in unterschiedlichen Wissensdomänen verstärkt. So gehen Köller, Baumert und Schnabel (2001) davon aus, dass intra-individuelle Vergleichsprozesse der eigenen Fähigkeiten in unterschiedlichen Domänen maßgeblich die Persönlichkeits- und Interessensentwicklung beeinflussen.

Zusammengenommen kann dies auch den zunächst erwartungswidrigen Befund einer negativen Korrelation zwischen der mathematikspezifischen Fähigkeit und der Deutschnote erklären. Die höhere mathematikspezifische Fähigkeit geht mit einer gesteigerten mathematikspezifischen Lernmotivation einher, die die Auswahl der Lernaufgaben maßgeblich reguliert. Gleichzeitig ist die höhere mathematikspezifische Fähigkeit mit einer geringeren Lernmotivation für sprachliche Fächer assoziiert. Somit investieren Schüler mit hoher mathematikspezifischer Fähigkeit die zeitlichen und kognitiven Ressourcen auch primär in mathematische Lernaktivitäten und nicht in Lernaktivitäten, die mit dem Fach Deutsch verbunden sind. Folglich haben Schüler (gleicher allgemeiner kognitiver Fähigkeit) mit höherer mathematikspezifischer Fähigkeit tendenziell eine schlechtere Deutschnote.

## 10 Gesamtdiskussion

Im Rahmen der Gesamtdiskussion werden zunächst die zentralen Ergebnisse der vorliegenden Arbeit zusammengefasst (Abschnitt 10.1). Vor diesem Hintergrund wird der Forschungsansatz rekapituliert, Schülerleistungen und kognitive Fähigkeiten aus einer gemeinsamen Strukturperspektive (*coordinate measurement*, Messick, 1984) zu analysieren (Abschnitt 10.2). In diesem Zusammenhang werden mögliche Implikationen und Forschungsperspektiven für die folgenden Themengebiete erörtert:

- Psychologische Forschung (Abschnitt 10.3.1)
- Schülerleistungsstudien und Lehr-Lern-Forschung (Abschnitt 10.3.2)
- Pädagogische Praxis (Abschnitt 10.3.3)

In den jeweiligen Abschnitten werden Befunde aus der vorliegenden Arbeit einbezogen. Die einzelnen Ergebnisse aus den drei Teilstudien werden jedoch nicht nochmals diskutiert, denn sie wurden bereits in den Abschnitten zu den jeweiligen Forschungsfragen ausführlich behandelt.

### 10.1 Zusammenfassung

Das erste zentrale Ziel dieser Arbeit war, mathematische Schülerleistung als psychologisches Konstrukt zu spezifizieren. Das zweite zentrale Ziel war es dann, mathematische Schülerleistung aus dieser Perspektive besser zu verstehen. Die beiden Ziele wurden im Rahmen von drei Teilstudien auf Grundlage einer Stichprobe von Neuntklässlern aus der PISA-2000-Untersuchung anvisiert. Zur Spezifikation mathematischer Schülerleistung als psychologisches Konstrukt wurden in dieser Arbeit Strukturmodelle kognitiver Fähigkeiten als theoretisches Grundgerüst herangezogen (Studie 1). Ausgehend von den so konzeptionalisierten kognitiven Fähigkeiten mathematischer Schülerleistung wurde analysiert, wie sich Schüler unterschiedlicher Schulformen in den Fähigkeiten unterscheiden (Studie 2) und welche Beziehungen diese Fähigkeiten zu soziodemografischen und motivationalen Schülermerkmalen sowie Schulnoten aufweisen (Studie 3). Die zentralen Befunde der drei Teilstudien lassen sich wie folgt zusammenfassen.

*Studie 1.* In der ersten Teilstudie wurde mit konfirmatorischen Methoden untersucht, welche kognitiven Fähigkeiten interindividuelle Unterschiede in Maßen mathematischer Schülerleistung erklären können. Ausgehend von Strukturmodellen kognitiver Fähigkeiten und auf Basis empirischer Befunde wurden mehrere konkurrierende Modelle spezifiziert und empirisch geprüft. Gemeinsames Merkmal dieser Modelle war ihr hierarchischer Aufbau, der implizierte, dass eine globalere mathematische Fähigkeit spezifischeren mathematischen Fähigkeiten hierarchisch übergeordnet ist. Die zentralen Unterschiede zwischen den Modellen waren: erstens, ob eine generelle kognitive Fähigkeit im Modell integriert war. Zweitens, ob Maße mathematischer Schülerleistung von mehreren kognitiven Fähigkeiten gleichzeitig beeinflusst wurden. Drittens, ob spezifische mathematische Fähigkeiten nach operativen Gesichtspunkten (in Form von Arten mathematischen Arbeitens, Neubrand u.a., 2001) oder in Form mathematischer Stoffgebiete konzeptionalisiert wurden.

Die wichtigsten Ergebnisse der Modellprüfungen ergaben, dass eine stoffgebietsspezifische Klassifikation von Mathematikaufgaben die empirischen Relationen besser approximieren konnte als eine Klassifikation mathematischer Fähigkeiten nach operativen Gesichtspunkten. Weiterhin konnte ein Nested-Faktormodell, in dem mathematische Schülerleistung als Amalgam stoffge-

biesspezifischer Fähigkeiten, einer mathematikspezifischen Fähigkeit und der allgemeinen kognitiven Fähigkeit spezifiziert wurde, interindividuelle Unterschiede in den Maßen mathematischer Schülerleistung am besten erklären. Verbale Fähigkeiten (indiziert durch die Skalen des PISA-Lesetests) hatten keinen als bedeutsam zu interpretierenden Einfluss über diese Fähigkeiten hinaus. Auch das Standardmodell, das in nahezu allen bisherigen Studien zu mathematischer Schülerleistung verwendet wurde und eine konzeptionelle Trennung fluider Fähigkeit und genereller mathematischer Fähigkeit vorsieht, erklärte die empirischen Relationen zwischen den Indikatoren kognitiver Fähigkeiten gut, aber schlechter als das Nested-Faktormodell.

*Studie 2.* Vor dem Hintergrund, dass Schüler mathematische Fähigkeiten primär in der Schule erwerben (vgl. Abschnitt 2.3; Geary, 1995; Köller & Baumert, 2002) und der Besuch unterschiedlicher Schulformen einen Großteil der interindividuellen Unterschiede in mathematischen Schülerleistungstests erklärt (Baumert, Trautwein & Artelt, 2003), befasste sich die zweite Teilstudie mit Schulformunterschieden in den kognitiven Fähigkeiten mathematischer Schülerleistung. Dabei wurden zwei Teilaspekte analysiert: Niveau- und Heterogenitätsunterschiede in den kognitiven Fähigkeiten mathematischer Schülerleistung.

Hinsichtlich der Niveauunterschiede der kognitiven Fähigkeiten wurde erwartet, dass sich die Schulformen deutlich unterscheiden und die schulformspezifischen Mittelwerte folgende Rangfolge aufweisen: Hauptschule < Realschule < Gymnasium. Hierfür sprach (zusätzlich zu den empirischen Befunden früherer Studien), dass Schüler mit höherem Niveau kognitiver Fähigkeiten in der Regel auch höhere Schulformen besuchen (Baumert u.a., 2003, S. 283) und dass leistungsförderliche (vor allem kognitiv herausfordernde) Aspekte des Mathematikunterrichts an höheren Schulformen tendenziell stärker ausgeprägt sind (vgl. Abschnitt 5.4). Die wichtigsten Befunde waren, dass in allen drei untersuchten Schulformen (Hauptschule, Realschule und Gymnasium) die kognitiven Fähigkeiten mathematischer Schülerleistung sowohl in Form eines (skalarinvarianten) Standardmodells als auch in Form eines (skalarinvarianten) Nested-Faktormodells konzeptionalisiert werden konnten. In beiden Modellen zeigte sich auch das erwartete Befundmuster hinsichtlich der Schulformunterschiede der untersuchten kognitiven Fähigkeiten: Schulformspezifische Mittelwertunterschiede in der fluiden Fähigkeit und der allgemeinen kognitiven Fähigkeit waren substantiell und nahezu gleich groß. Auch für die generelle mathematische Fähigkeit und die mathematikspezifische Fähigkeit resultierten deutliche Schulformunterschiede, die jedoch bei der generellen mathematischen Fähigkeit wesentlich stärker ausgeprägt waren als bei der mathematikspezifischen Fähigkeit.

Die Analysen zur Heterogenität der kognitiven Fähigkeiten mathematischer Schülerleistung konzentrierten sich auf die mathematikspezifische Fähigkeit. Ausgangspunkt zur Ableitung von einer generellen und einer schulformspezifischen Differenzierungshypothese waren die empirischen Befunde früherer Studien, dass Personen mit höherem Fähigkeitsniveau eine stärkere Differenzierung der kognitiven Fähigkeiten aufweisen. Die Ergebnisse in der vorliegenden Arbeit stützten die generelle Differenzierungshypothese, da die Heterogenität der mathematikspezifischen Fähigkeit in der Gesamtpopulation der Schüler mit zunehmendem Fähigkeitsniveau zunahm. Eine mögliche Erklärung dieser Differenzierung ist, dass Schüler mit höherem Fähigkeitsniveau Zugang zu besseren Bildungsmöglichkeiten haben und somit bessere Investitionsmöglichkeiten für ihre allgemeine kognitive Fähigkeit vorfinden. Entgegen der Erwartung konnte diese schulformspezifische Differenzierungshypothese nur für Schülergruppen im jeweils unteren schulformspezifischen Fähigkeitspektrum bestätigt werden. Im oberen schulformspezifischen Fähigkeitspektrum war die Heterogenität der mathematikspezifischen Fähigkeit weitestgehend unabhängig von der besuchten Schulform.



*Studie 3.* Über die Schulformunterschiede hinaus sollte ein detailliertes Bild der kognitiven Fähigkeiten mathematischer Schülerleistung mit Blick auf ihre externe Validität zu mehreren soziodemografischen und motivationalen Schülermerkmalen sowie Schulnoten gezeichnet werden.

Wie auf Grundlage von Befunden aus früheren Studien erwartet wurde, waren im Vergleich zur generellen mathematischen Fähigkeit Geschlechterunterschiede (zu Gunsten der Jungen) in der mathematikspezifischen Fähigkeit deutlich stärker ausgeprägt. Es bestanden nahezu keine Geschlechterunterschiede hinsichtlich der allgemeinen kognitiven Fähigkeit und der fluiden Fähigkeit.

Weiterhin waren – wie auch bereits frühere Studien zeigten – Unterschiede des familiären Hintergrunds (indiziert durch die Anzahl an Büchern und den sozioökonomischen Status) positiv mit der generellen mathematischen Fähigkeit, der fluiden Fähigkeit und der allgemeinen kognitiven Fähigkeit korreliert. Hingegen bestand nahezu kein Zusammenhang zwischen dem familiären Hintergrund und der mathematikspezifischen Fähigkeit.

Mit Blick auf die externe Validität zu motivationalen Schülermerkmalen wurden das Interesse und das Selbstkonzept untersucht. Hinsichtlich des mathematischen und verbalen Selbstkonzepts wurden Hypothesen auf Grundlage des Bezugsrahmenmodells (Marsh, 1986) abgeleitet. Die Ergebnisse stützen alle Modellannahmen: Schüler mit stärker ausgeprägten mathematischen Fähigkeiten hatten (bei Kontrolle der generellen verbalen Fähigkeit bzw. der spezifischen verbalen Fähigkeit sowie der fluiden Fähigkeit bzw. der allgemeinen kognitiven Fähigkeit) ein positiveres mathematisches Selbstkonzept von sich. Die mathematischen Fähigkeiten hatten darüber hinaus jeweils einen negativen Effekt auf das verbale Selbstkonzept. Weiterhin hatte im Regressionsmodell die allgemeine kognitive Fähigkeit jeweils positive Effekte auf die fachspezifischen Selbstkonzepte, während die fluide Fähigkeit die Selbstkonzepte kaum beeinflusste.

Ausgehend von der Überlegung, dass die Interessensgenese vom Kompetenzerleben abhängig ist, wurde ein erweitertes Bezugsrahmenmodell geprüft. Es zeigte sich wie erwartet, dass stärker ausgeprägte Selbstkonzepte mit einem höheren korrespondierenden fachspezifischen Interesse assoziiert waren. Erwartungswidrig hatte das mathematische Selbstkonzept bei Kontrolle des verbalen Selbstkonzepts keinen negativen Effekt auf das Interesse am Lesen. Ebenso hatte das verbale Selbstkonzept keinen bedeutsamen negativen Effekt auf das mathematische Interesse. Über die Selbstkonzepte hinaus konnten die kognitiven Fähigkeiten das mathematische Interesse nur geringfügig, dafür aber das Interesse am Lesen substantiell vorhersagen: Bei Kontrolle des verbalen und des mathematischen Selbstkonzepts waren Schüler mit höheren Leistungen in der fluiden Fähigkeit, der generellen verbalen Fähigkeit und in der allgemeinen kognitiven Fähigkeit mehr am Lesen interessiert. Hingegen hatten hierbei Schüler mit stärker ausgeprägter mathematikspezifischer Fähigkeit oder einer höheren Leistung in der generellen mathematischen Fähigkeit ein geringeres Interesse am Lesen.

Wie auf Grundlage der Modelle schulischen Lernens erwartet wurde, erzielten Schüler mit höheren Leistungen in der generellen mathematischen Fähigkeit, in der fluiden Fähigkeit, in der mathematikspezifischen Fähigkeit und in der allgemeinen kognitiven Fähigkeit bessere Mathematiknoten. Im Gegensatz zu den drei anderen kognitiven Fähigkeiten (diese korrelierten jeweils positiv mit der Deutschnote) hatten jedoch Schüler mit höheren Leistungen in der mathematikspezifischen Fähigkeit schlechtere Noten in Deutsch.

## 10.2 Schülerleistung versus kognitive Fähigkeiten? Ein Fazit

Zusammenfassend betrachtet belegten die drei Teilstudien die theoretische und empirische Tragfähigkeit eines *coordinate measurement* von mathematischer Schülerleistung und Indikatoren kognitiver Fähigkeiten, wie es Samuel Messick (1984) anregte: Die didaktisch-spezifizierte „mathematische Schülerleistung“ wurde in psychologische Strukturmodelle kognitiver Fähigkeiten integriert und somit in Form psychologischer Konstrukte spezifiziert.

Der gewählte Ansatz implizierte, Grenzen zwischen Forschungsdisziplinen zu überschreiten. So wurden in dieser Arbeit theoretische Überlegungen aus der differenziellen und pädagogischen Psychologie, aus der Mathematikdidaktik und der pädagogisch orientierten Lehr-Lern-Forschung integriert, um die jeweiligen Forschungsfragen zu motivieren und Hypothesen abzuleiten.

Ein *coordinate measurement* kann jedoch auch mit „Kosten“ verbunden sein. Zum Beispiel wurde mit der Konzeptualisierung der kognitiven Fähigkeiten mathematischer Schülerleistung in Form des Nested-Faktormodells ein aus Sicht vieler pädagogischer Forscher eher heikles Konstrukt (siehe hierzu Plomin & Petrill, 1997) wieder „rehabilitiert“: die allgemeine kognitive Fähigkeit (bzw. die psychometrisch definierte „Intelligenz“). Diese erklärte zusätzlich zur mathematikspezifischen Fähigkeit und stoffgebietsspezifischen Fähigkeiten substanziell interindividuelle Unterschiede bei Maßen mathematischer Schülerleistung.

Unter anderem um diese „Kosten“ zu rechtfertigen, wurden Schulformunterschiede und die externe Validität der kognitiven Fähigkeiten mathematischer Schülerleistung sowohl bei einer Konzeptualisierung in Form des Standardmodells als auch in Form des Nested-Faktormodells untersucht. Damit konnten zwei Ziele simultan verfolgt werden:

- Bei Verwendung des Standardmodells konnten die Befunde aus früheren Studien repliziert werden.
- Durch die Verwendung des Nested-Faktormodells wurde ein gegenüber dem Standardmodell detaillierteres Bild hinsichtlich von Schulformunterschieden und der externen Validität der kognitiven Fähigkeiten mathematischer Schülerleistung gezeichnet.

Insbesondere die Befunde für das Nested-Faktormodell zeigten, dass es sich lohnt, die Perspektive eines *coordinate measurement* von Schülerleistungen und kognitiven Fähigkeiten einzunehmen und die Trennung der Konstrukte (vgl. Abschnitt 4.1) ernsthaft in Frage zu stellen. Theoretische und empirische Argumente (siehe Kap. 7, 8 und 9) lieferte die vorliegende Arbeit. Um die Tragfähigkeit dieses Ansatzes weiter zu elaborieren, werden in den nächsten Abschnitten mögliche Implikationen und Perspektiven für die psychologische Forschung, für Schülerleistungsstudien und die Lehr-Lern-Forschung sowie für die pädagogische Praxis erörtert. Im Zentrum steht hierbei die Konzeptualisierung der kognitiven Fähigkeiten mathematischer Schülerleistung in Form des Nested-Faktormodells.

## 10.3 Implikationen und Perspektiven

### 10.3.1 Psychologische Forschung

Mögliche Implikationen und Perspektiven für die psychologische Forschung werden anhand von zwei Themen diskutiert: alternative Konzeptualisierungen der kognitiven Fähigkeiten mathematischer Schülerleistung und der Brückenschlag zwischen psychologischen Forschungsdisziplinen.

*Alternative Konzeptualisierungen.* In dieser Arbeit wurden alternative Strukturmodelle mathematischer Schülerleistung miteinander verglichen. Im Forschungskontext kann insgesamt gesehen eine Konzeptualisierung in Form des Nested-Faktormodells favorisiert werden. Drei Gründe sprachen dafür:

- Eine Spezifikation der kognitiven Fähigkeiten mathematischer Schülerleistung in Form des Nested-Faktormodells konnte die empirischen Relationen im Vergleich zu den rivalisierenden Modellen am besten approximieren (vgl. Abschnitt 7.2).
- Mithilfe des Nested-Faktormodells konnten Schulformunterschiede in der mathematikspezifischen Fähigkeit und der allgemeinen kognitiven Fähigkeit unabhängig voneinander messfehlerbereinigt analysiert werden (vgl. Kap. 8).
- Mit dem Nested-Faktormodell war es möglich, die bivariaten, aber auch die multivariaten Beziehungen zu einigen zentralen Außenkriterien wesentlich differenzierter zu beleuchten, als dies mit dem Standardmodell der Fall war (vgl. Kap. 9).

Allerdings sind bei einer Verwendung des Nested-Faktormodells nicht nur Vorzüge verbunden, und es ist auch möglich ähnliche Befundmuster durch alternative Analysestrategien zu erhalten. Letzteres machten auf den ersten Blick insbesondere die Analysen zum Bezugsrahmenmodell (Abschnitt 9.3) und zum erweiterten Bezugsrahmenmodell (Abschnitt 9.4) deutlich. Für die domänenspezifischen Fähigkeiten ( $M$  und  $M'$  bzw.  $V$  und  $V'$ ) war das Befundmuster für das Standardmodell und das Nested-Faktormodell äquivalent. Jedoch wurde bei Verwendung des Standardmodells „verschleiert“, welche Varianzanteile bei den Selbstkonzepten durch welche Fähigkeit erklärt wurden. So resultierte beispielsweise trotz positiver bivariater Korrelationen der fluiden Fähigkeit mit den fachspezifischen Selbstkonzepten aufgrund der Kolinearität mit den beiden anderen Fähigkeiten ein Effekt um Null. Im Gegensatz hierzu beeinflusste im Nested-Faktormodell die allgemeine kognitive Fähigkeit die fachspezifischen Selbstkonzepte substanziell positiv. Der Grund für diese zunächst widersprüchlichen Befunde war, dass die generelle verbale Fähigkeit und die generelle mathematische Fähigkeit jeweils die gemeinsamen Varianzanteile von fluider Fähigkeit und den fachspezifischen Selbstkonzepten erklärten. Weiterhin konnten die negativen Effekte auf die Selbstkonzepte der jeweils anderen Domäne unter anderem deshalb gefunden werden, weil bei einer regressionsanalytischen Kontrolle auch gleichzeitig für die fluide Fähigkeit kontrolliert wurde.

Der meines Erachtens große Vorteil des Nested-Faktormodells ist also, dass durch die explizite Konzeptualisierung der kognitiven Fähigkeiten und aufgrund der wechselseitigen Orthogonalität dieser Fähigkeiten die Effekte in Regressionsmodellen unmittelbar auf die jeweiligen Fähigkeiten attribuiert werden können.

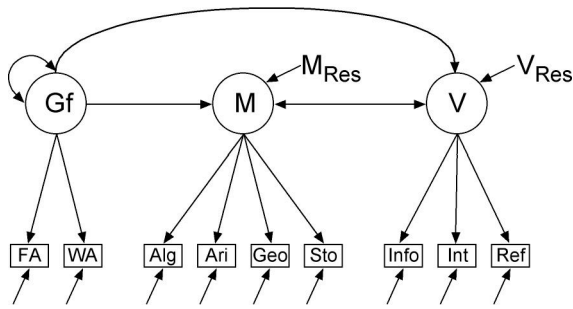


Abbildung 31: Standardmodell: Regressionsanalytische Kontrolle für die fluide Fähigkeit

Das Nested-Faktormodell ist auch im Vergleich zu einer regressionsanalytischen Kontrolle der fluiden Fähigkeit in der generellen mathematischen Fähigkeit (oder generellen verbalen Fähigkeit) vorteilhaft (Abb. 31).

Die regressionsanalytische Kontrolle im Standardmodell impliziert wie auch im *Higher-Order-Modell* eine Proportionalitätsrestriktion (siehe Abschnitt 9.1.1). Beispielsweise wird der Einfluss der fluiden Fähigkeit wie auch der residualisierten generellen mathematischen Fähigkeit ( $M_{Res}$ ) auf die Maße mathematischer Schülerleistung beide Male über die gleichen Faktorladungen vermittelt. Damit sind  $M_{Res}$  und die mathematikspezifische Fähigkeit nicht äquivalent, da Letztere unabhängig von der allgemeinen kognitiven Fähigkeit die Maße mathematischer Schülerleistung beeinflussen kann. Um also die Beziehungen zwischen Außenkriterien und einer globaleren mathematischen Fähigkeit unabhängig von der fluiden Fähigkeit (bzw. der allgemeinen kognitiven Fähigkeit) möglichst exakt zu bestimmen, ist das Nested-Faktormodell das Modell der Wahl (siehe auch Schmiedek, 2003).

Eine offene Frage ist, ob das Nested-Faktormodell für Menschen aller Altersstufen spezifiziert werden kann. Folgt man Geary (1994, 1996), dann kann erst für Jugendliche an der Highschool ein Faktor extrahiert werden, der die Fähigkeit zum mathematischen Problemlösen repräsentiert. Dies würde (bei der Interpretation der mathematikspezifischen Fähigkeit als Fähigkeit zum mathematischen Problemlösen) das Problem implizieren, dass eine Konzeptualisierung der kognitiven Fähigkeiten mathematischer Schülerleistung in Form des Nested-Faktormodells erst für Jugendliche in der Mitte oder am Ende der Sekundarstufe I möglich ist. In Anlehnung an Geary scheint für Jugendliche zu diesem Zeitpunkt die Differenzierung der mathematikspezifischen Fähigkeit so weit fortgeschritten, dass ein korrespondierender Faktor empirisch identifiziert ist. Eine Überprüfung dieser entwicklungspsychologischen Hypothese könnte eine Aufgabe für zukünftige Forschung sein.

Ein weiterer Nachteil bei der Konzeptualisierung der kognitiven Fähigkeiten mathematischer Schülerleistung in Form des Nested-Faktormodells liegt in der Ambiguität der Operationalisierungen (Schulze, 2005). Bei einer Konzeptualisierung in Form des Standardmodells werden die jeweiligen kognitiven Fähigkeiten *eindeutig* operationalisiert. Maße mathematischer Schülerleistung sind somit nur Indikatoren der generellen mathematischen Fähigkeit. Hingegen scheint mit herkömmlichen Maßen mathematischer Schülerleistung eine eindeutige Operationalisierung der mathematikspezifischen Fähigkeit nicht möglich zu sein. Im Nested-Faktormodell indizieren diese manifesten Variablen die mathematikspezifische Fähigkeit und gleichzeitig die allgemeine kognitive Fähigkeit. Es stellt sich die Frage, ob es möglich ist, Items (oder Erhebungsmethoden) zu generieren, die nur die mathematikspezifische Fähigkeit erfassen. Hierbei ist aber einschrän-

kend festzustellen, dass generell in der Psychologie die Eindimensionalität der Messungen zwar angestrebt, aber selten erreicht wird. Bollen und Lennox (1991) konstatieren zum Beispiel: „exploratory factor analysis virtually always leads to each indicator having nonzero loadings on more than one factor“ (Bollen & Lennox, 1991, S. 309).

Da die mathematikspezifische Fähigkeit nicht eindeutig operationalisiert werden kann und somit zunächst ein statistisches „Phänomen“ ist, ist ihre Interpretation ebenfalls nicht eindeutig. Damit verbundene Einschränkungen und Perspektiven werden im nächsten Abschnitt noch weiter ausgearbeitet.

*Brückenschlag zwischen Forschungsdisziplinen.* In dieser Arbeit wurden die kognitiven Fähigkeiten mathematischer Schülerleistung, die in Form des Nested-Faktormodells konzeptualisiert wurden, wie folgt interpretiert: Die stoffgebietsspezifischen Residualterme repräsentieren die Verfügbarkeit und Fähigkeit zur Anwendung stoffgebietsspezifischen Wissens. Die mathematikspezifische Fähigkeit repräsentierte die stoffgebietsunabhängige Fähigkeit zum mathematischen Problemlösen, und die allgemeine kognitive Fähigkeit repräsentierte die Arbeitsgedächtniskapazität.

Wichtig ist bei diesen Interpretationen zu bedenken, dass sie nicht empirisch durch operationalisierte Messungen der kognitiven Prozesse oder Wissensarten gestützt wurden. Hier ist auch noch in dreierlei Hinsicht ein Defizit in der Forschung festzustellen:

Erstens hat es bisher nur sehr wenige Studien gegeben (Geary & Widaman, 1992; Swanson & Beebe-Frankenberger, 2004; Tirre & Pena, 1993), die explizit den Zusammenhang zwischen interindividuellen Unterschieden bei kognitiven Informationsverarbeitungsprozessen und Wissensarten und Maßen mathematischer Schülerleistung oder Maßen mathematischer Fähigkeiten untersuchten. Oftmals stehen bei Studien, die beispielsweise den Zusammenhang zwischen Arbeitsgedächtniskapazität und schlussfolgerndem Denken analysieren, mathematische Fähigkeiten nicht im Mittelpunkt (Ackerman, Beier & Boyle, 2002; Kyllonen & Christal, 1990; Süß, Oberauer, Wittmann, Wilhelm & Schulze, 2002) oder werden gar nicht erst berücksichtigt (Engle, Tuholski, Laughlin & Conway, 1999).

Zweitens ist es generell schwierig, Wissensarten *eindeutig* zu operationalisieren. Die Analyse mathematischer Schülerleistung oder mathematischer Fähigkeiten in dieser Arbeit bildete hier keine Ausnahme. Unabhängig davon, ob man Mathematikaufgaben des PISA-Tests inhaltlich oder operativ klassifizierte, zeigten die hohen latenten Interkorrelationen (siehe Abschnitt 7.1.3), dass die korrespondierenden Faktoren nur zu relativ geringen Teilen *unique Anteile* der jeweiligen Fähigkeiten bzw. stoffgebietsspezifischen Wissensarten messen.

Ein weiteres Problem bei der Operationalisierung von mathematischem Wissen ist die *Nominalistic Fallacy* (Cliff, 1983, S. 120), die pointiert besagt, dass die alleinige Benennung einer Aufgabe als Indikator für Wissen dazu führt, dass diese Aufgabe Wissen erfasst. Vor dieser *Fallacy* sind selbst die Autoren der methodisch besten Studien (z.B. Tirre & Pena, 1993) nicht gefeit. So haben Tirre und Pena die Aufgabe *decomposition and sequencing* als Indikator für mathematisches Wissen gehandelt. Allerdings wird die gleiche (oder eine nahezu identische) Aufgabe in der Strukturforschung kognitiver Fähigkeiten unter dem Namen *Necessary Arithmetic Operations* als Indikator für quantitatives Reasoning (Carroll, 1993) oder allgemeine kognitive Fähigkeit (Snow & Lohman, 1989) betrachtet. Möchte man den Einfluss von Wissen und kognitiven Fähigkeiten getrennt untersuchen, dann sollte bei der Benennung der jeweiligen Maße sehr viel Sorgfalt walten und die relevante Literatur mit Umsicht rezipiert werden.

Drittens gibt es auch keine Studien, die explizit versuchen, „faktoriell reine“ Maße mathematischer Problemlöseprozesse, wie sie Mayer (1985) unterscheidet, zu entwickeln. Wohl auch deshalb, weil die Erfassung von Prozessparametern der Informationsverarbeitung ein schwieriges Unterfangen ist (Lohman, 1994).

Insgesamt gesehen gibt es auf dem Forschungsfeld zum Brückenschlag zwischen psychometrisch definierten kognitiven Fähigkeiten, der Wissenspsychologie und kognitionspsychologischen Informationsverarbeitungsprozessen noch sehr viel zu tun, oder um mit Geary zu sprechen:

Individual-difference studies that explicitly examine these skills [the ability to mentally translate and represent the meaning of arithmetical and algebraic word problems, ease with which the associated schemas develop], in concert with arithmetical processing and working memory skills, for their relation to performance on mathematical reasoning tests are needed to fill in the gaps in our understanding of this area. (Geary, 1994, S. 147)

Über den Brückenschlag zwischen Psychometrie, Wissens- und Kognitionspsychologie hinaus eröffnet die Konzeption der kognitiven Fähigkeiten mathematischer Schülerleistung in Form des Nested-Faktormodells möglicherweise auch interessante Analysemöglichkeiten für die Verhaltensgenetik. So konvergieren die Befunde der meisten verhaltensgenetischen Studien dahingehend, dass in westlichen industrialisierten Gesellschaften die Heritabilität<sup>30</sup> der allgemeinen kognitiven Fähigkeit bei ungefähr 50 Prozent liegt (Plomin, DeFries, McClearn & Mc Guffin, 2001; Weinert, 1994). In ähnlicher Höhe liegt auch die Heritabilität der (über mehrere Fächer gemittelten) allgemeinen Schülerleistung (Kovas, Harlaar, Petrill & Plomin, 2005; Plomin u.a., 2001; Plomin & Petrill, 1997). Hinsichtlich von Leistungsindikatoren (Tests und Lehrerurteile) der mathematischen Schülerleistung liegen die Heritabilitätsschätzungen zwischen .20 und .66 (Plomin u.a., 2001). Die Heritabilität der Fähigkeit zum mathematischen Problemlösen (Geary, 1994, verwendet hierbei den Begriff „quantitatives Reasoning“) liegt im Bereich von .27 und .66.

Ist mathematische Schülerleistung (insbesondere die Fähigkeit zum mathematischen Problemlösen) nun „vererbbar“? Nach Geary (1994, S. 152–153) ist bei den Heritabilitäten zur Fähigkeit zum mathematischen Problemlösen zu bedenken, dass die geschätzten Koeffizienten in erster Linie auf die allgemeine kognitive Fähigkeit (Geary spricht hierbei von generellem Reasoning) zurückzuführen sind und nicht auf die spezifische Fähigkeit zum mathematischen Problemlösen attribuiert werden können. Denn diese Fähigkeit entsteht nach Geary erst durch die intensive Auseinandersetzung mit mathematischen Problemen. Gearys Überlegung sollte auch auf die Heritabilitäten für die mathematische Schülerleistung übertragbar sein, da in die Varianz der mathematischen Schülerleistung – entsprechend den Modellen schulischen Lernens (Abschnitt 5.1) und der Ergebnisse dieser Arbeit (siehe Abb. 17) – auch bedeutsame Varianzanteile der allgemeinen kognitiven Fähigkeit eingehen.

Folgt man dieser Argumentation von Geary, dann kann spekuliert werden, dass die allgemeine kognitive Fähigkeit in dieser Arbeit eine deutlich höhere Heritabilität aufweist als die mathematikspezifische Fähigkeit. Allerdings kann diese Vermutung nicht anhand der Daten in dieser Studie gestützt werden, da die Varianzanteile, die durch Umweltunterschiede oder genetische

---

<sup>30</sup> Die Heritabilität gibt an, welcher Anteil der phänotypischen Varianz eines Persönlichkeitsmerkmals (z.B. die allgemeine kognitive Fähigkeit) durch die genetischen Unterschiede, die zwischen Personen bestehen, erklärt werden kann. Dabei wird angenommen, dass der Anteil der Varianz des Persönlichkeitsmerkmals, der nicht durch die genetischen Unterschiede erklärt wird, durch Unterschiede in den Umwelteinflüssen (und Messfehler) bedingt wird. Wichtig bei der Interpretation der Heritabilität ist zu beachten, dass sich diese stets auf eine bestimmte Population zu einem bestimmten Zeitpunkt bezieht (Plomin u.a., 2001).

Unterschiede erklärt werden, nicht bestimmt werden können. Dies ist jedoch eine lohnenswerte Frage für zukünftige Forschungsarbeiten, die auch die Tragfähigkeit der Konzeptualisierung kognitiver Fähigkeiten in Form des Nested-Faktormodells in der Verhaltensgenetik prüft (siehe auch Schmiedek, 2003).

Unabhängig von der exakten Bestimmung der Heritabilitätskoeffizienten kognitiver Fähigkeiten ist die weit wichtigere Frage zur Erbe-Umwelt-Diskussion zu erklären, wie genetische Faktoren und Umweltfaktoren zusammenwirken (z.B. Bronfenbrenner & Ceci, 1994). Denn wie sich eine genetische Disposition entwickelt, wird zu großen Teilen von der Umwelt mitbestimmt (Gottfredson, 1997; Neisser u.a., 1996; Snyderman & Rothman, 1987). Oder um mit Plomin und Kollegen zu sprechen: „For a single individual, both genotype and environment are indispensable – a person would not exist without both genes and environment.“ (Plomin u.a., 2001, S. 87)

Angesichts der Tatsache, dass mathematische Schülerleistung primär das Ergebnis schulischer Lernprozesse (Geary, 1995; Köller & Baumert, 2002) und damit abhängig von Umweltfaktoren ist, lohnt es sich, mögliche Implikationen und Perspektiven anhand der Ergebnisse der vorliegenden Arbeit für Schülerleistungsstudien und die Lehr-Lern-Forschung zu erörtern.

### 10.3.2 Schülerleistungsstudien und Lehr-Lern-Forschung

*Schülerleistungsstudien.* Ein Leitgedanke für Schülerleistungsstudien und die Lehr-Lern-Forschung ist, dass die eingesetzten Maße mathematischer Schülerleistung die intendierten Konstrukte abbilden (American Educational Research Association, 1999; Millman & Greene, 1989). Aus methodischer und konzeptioneller Sicht ist hierbei das Nested-Faktormodell sehr interessant, denn es bot in dieser Arbeit die Möglichkeit, den Einfluss der allgemeinen kognitiven Fähigkeit, der mathematikspezifischen Fähigkeit und stoffgebietsspezifischer Fähigkeiten getrennt voneinander zu analysieren.

Einschränkend ist mit Blick auf die Ergebnisse einzuräumen, dass die stoffgebietsspezifischen Fähigkeiten (wie auch spezifische operative Fähigkeiten) nicht explizit modelliert werden konnten (die korrespondierenden Varianzanteile wurden in Abschnitt 7.2.3 nur indirekt über die Reliabilitäten der WLE-Scores berechnet). Der ausschlaggebende Grund hierfür war, dass die Schüler nicht genügend Mathematikaufgaben der jeweiligen Stoffgebiete bearbeitet hatten, damit zwei oder mehr Indikatoren für das jeweilige Stoffgebiet geschätzt werden konnten. Denn nur bei Vorliegen mehrerer manifesten Variablen pro Stoffgebiet hätten die stoffgebietsspezifischen Fähigkeiten direkt spezifiziert werden können.

Angesichts der Einschränkungen wäre eine bedeutsame Erweiterung des Nested-Faktormodells, wie es in der vorliegenden Arbeit verwendet wurde, stoffgebietsspezifische und spezifische operative Fähigkeiten sowie deren Kombinationen (z.B. rechnerische Modellierungsaufgaben aus der Arithmetik) als wechselseitig unabhängige Fähigkeiten explizit zu modellieren. Allerdings müssten Testbatterien zur Erfassung von Schülerleistungen dann idealerweise zwei Anforderungen erfüllen:

- Es sollten mindestens zwei verschiedene Maße fluider Fähigkeiten (z.B. Figurenanalogien und Wortanalogien) von jedem Schüler bearbeitet werden. Damit ist es möglich, die allgemeine kognitive Fähigkeit und die mathematikspezifische Fähigkeit analog zu Modell 7a (Abb. 16) zu

spezifizieren. Sollte es (z.B. aufgrund der begrenzten Testzeit) nur möglich sein, maximal einen Indikator der fluiden Fähigkeit einzusetzen, ist in Anlehnung an Wilhelm (2005) eine figurale Reasoningaufgabe wie die Figurenanalogien empfehlenswert.

- Um Aussagen über operative Fähigkeiten, stoffgebietsspezifische Fähigkeiten sowie deren Kombinationen treffen zu können, müssen die Schüler genügend korrespondierende Mathematikaufgaben bearbeiten, damit mindestens zwei reliable Indikatoren dieser Fähigkeiten sowie der Fähigkeitskombinationen ermittelt werden können.

Prinzipiell gilt für die Erfassung latenter Variablen, dass mit zunehmender Anzahl bearbeiteter Items die Reliabilität der korrespondierenden WLE-Scores zunimmt. Daher ist eine möglichst große Anzahl bearbeiteter Items wünschenswert. Aber natürlich ist dabei zu bedenken, dass die Testzeit insgesamt begrenzt ist und somit die Zumutbarkeit bzw. die Belastung der Schüler durch die Bearbeitung einer großen Leistungstestbatterie in Rechnung zu stellen ist (Baumert, Stanat & Demmrich, 2001). Eine Möglichkeit, sowohl Aussagen über ein breites Kompetenzspektrum zu machen wie auch die Testbelastung in Grenzen zu halten, bietet dann das Multi-Matrix-Testadministrationsdesign (Mislevy u.a., 1992).

Eine weitere Möglichkeit zur Begrenzung der Anzahl bearbeiteter Items bieten möglicherweise auch Nested-Faktormodelle, die auf Itemebene als Item-Response-Modelle spezifiziert werden (Wu, Adams & Wilson, 1998). Allerdings sind die Eigenschaften dieser Modelle im Gegensatz zu den Modellen der konfirmatorischen Faktorenanalyse, die auf der Ebene der WLE-Scores ansetzen, kaum untersucht. Weiterhin ist dabei auch die Tatsache in Rechnung zu stellen, dass (ausgehend von den Modellen 1 bis 3, Abb. 14) die benötigte Rechenzeit (bei einem Pentium(R) 4, 2.8 GHz und Daten von 29.386 Schülern) für die Modellanalysen bei diesen Modellen bei mehreren Tagen liegen kann, wohingegen die benötigte Rechenzeit bei den Nested-Faktormodellen, wie sie in der vorliegenden Arbeit spezifiziert wurden, bei wenigen Sekunden liegt. Bei einer Vielzahl untersuchter Modelle ist dies – wie in der vorliegenden Arbeit – ein nicht zu unterschätzender zeitökonomischer Aspekt.

Bemerkenswert ist allerdings, dass die Nested-Faktormodelle auf Itemebene auch dahingehend erweiterbar sind, dass simultan die Beziehungen zu verschiedenen Außenkriterien analysiert werden können. Somit stellen die IRT-Modelle eine interessante Analyseperspektive dar, wenn die Computerleistungen weiter zunehmen. Bis es soweit ist, ist es empfehlenswert, multidimensionale Raschmodelle mit interkorrelierten Gruppenfaktoren und darauf aufbauende Modelle (siehe z.B. Abb. 26 in Abschnitt 9.1.2, Abb. 28 in Abschnitt 9.3 oder Abb. 29 in Abschnitt 9.4) zu spezifizieren. Wichtig ist dabei, dass immer jeweils minimal zwei latente Variablen pro Stoffgebiet, operative Fähigkeit oder die jeweilige Fähigkeitskombination spezifiziert werden. Eine ausreichende Itemanzahl vorausgesetzt, können dann zwei oder mehr WLE-Scores (siehe Rost, 2004a zu alternativen Personenparametern) pro Fähigkeitsdimension geschätzt und darauf basierende Nested-Faktormodelle bzw. hierzu alternative Konzeptualisierungen der kognitiven Fähigkeiten mathematischer Schülerleistung analysiert werden.

*Lehr-Lern-Forschung.* Mit der getrennten Konzeptualisierung der allgemeinen kognitiven Fähigkeit und davon unabhängiger mathematischer Fähigkeiten eröffnet das Nested-Faktormodell auch aufschlussreiche Perspektiven für die Lehr-Lern-Forschung (für methodische Anforderungen an diese Studien siehe Abschnitt 8.3). Mit diesem Modell können Effekte des Mathematikunterrichts auf alle spezifizierten Fähigkeiten simultan untersucht werden. Hiermit kann insbesondere der Frage nachgegangen werden, ob der Mathematikunterricht primär auf die mathema-



tischen Fähigkeiten wirkt, oder ob nicht – entgegen der Vermutung Cattells (1987) – auch die fluide Fähigkeit bzw. die allgemeine kognitive Fähigkeit positiv beeinflusst werden.

Die umfangreichste Zusammenfassung empirischer Studien zu diesem Thema stammt von Ceci (1991). Gemeinsames Merkmal der Studien dieser Literaturübersicht ist, dass Stichproben untersucht wurden, die sich möglichst nur im Ausmaß des Schulbesuchs unterscheiden. Auf Grundlage aller analysierten Studien kommt Ceci (1991, S. 717) in seiner Arbeit zu dem Schluss, dass ein Jahr Schule einen positiven Effekt auf die allgemeine kognitive Fähigkeit hat, der bis zu .40 Standardabweichungen (das sind 6 IQ-Punkte bei einer Standardabweichung von 15 IQ-Punkten) betragen kann.

Der Schulbesuch beeinflusst aber nicht nur die allgemeine kognitive Fähigkeit. Mit Blick auf mathematische Fähigkeiten zeigten Studien aus der Lehr-Lern-Forschung (vgl. Abschnitt 5.3 und 5.4), dass Unterschiede in den Basisdimensionen der Unterrichtsqualität, in der Qualität des Übens (repetitiv vs. anspruchsvoll z.B. in variierenden Kontexten) und in kognitiv herausfordernden Unterrichtselementen einen positiven Effekt auf die Leistungsentwicklung in Mathematik haben.

Unter der Annahme, dass leistungsförderliche Aspekte des Mathematikunterrichts in erster Linie mathematische Fähigkeiten fördern, sind die Befunde aus der vorliegenden Arbeit relevant: Die mathematikspezifische Fähigkeit ist (weitestgehend) unabhängig vom familiären Hintergrund eines Schülers (vgl. Abschnitt 9.2) und sollte daher primär das Ergebnis schulischer Lehr-Lern-Prozesse sein (Geary, 1995; Köller & Baumert, 2002). Daher sind auch die größten Wirkungen des Mathematikunterrichts bei der mathematikspezifischen Fähigkeit zu erwarten. Es könnte sein, dass aufgrund der bisherigen Dominanz des Standardmodells in der Lehr-Lern-Forschung die Wirkungen des Mathematikunterrichts unterschätzt worden sind. Eine Konzeptionalisierung der kognitiven Fähigkeiten mathematischer Schülerleistung in Form eines Nested-Faktormodells würde es ermöglichen, die differenzielle Wirkung des Mathematikunterrichts getrennt für die allgemeine kognitive Fähigkeit und die mathematikspezifische Fähigkeit zu untersuchen und dieser Annahme gezielt nachzugehen.

Über die Beantwortung dieser Frage hinaus könnte die Verwendung des Nested-Faktormodells auch für die Analyse von *Aptitude-Treatment*-Interaktionen (ATI) mit verschiedenen Unterrichtsformen lohnenswert sein. Die zentrale Annahme von ATI besagt, dass nicht alle Schüler gleichermaßen von bestimmten Unterrichtsformen profitieren. Ziel des ATI-Forschungsprogramms ist es, für jeden Schüler entsprechend seiner individuellen Charakteristika (A) die Lernform (T) zu identifizieren, bei der er die Zielkriterien des Unterrichts bestmöglich erreicht. Eine erste umfassende Zwischenbilanz des ATI-Forschungsprogramms zogen Cronbach und Snow (1977). Darin kommen sie zu dem Schluss, dass erwartungswidrig häufiger Interaktionen zwischen *treatment* und der allgemeinen kognitiven Fähigkeit gefunden wurden als mit spezifischen Fähigkeiten.

Dieser Befund ist verständlich, wenn man die Ergebnisse der vorliegenden Arbeit heranzieht. In den Studien zu ATI wurden in der Regel nur messfehlerbehaftete Skalenscores verwendet. Es ist davon auszugehen, dass diese in erster Linie Varianzanteile der allgemeinen kognitiven Fähigkeit und nur zu geringen Anteilen Varianzanteile inhaltsgebundener Fähigkeiten enthielten. Beispielsweise gingen bei den stoffgebietsspezifischen WLE-Scores in der vorliegenden Arbeit maximal 15 Prozent der Gesamtvarianz des Arithmetikscores auf die mathematikspezifische Fähigkeit, aber bis zu 51 Prozent der Gesamtvarianz des Algebrascors auf die allgemeine kognitive Fähigkeit zurück. Dieses Ergebnis ist kein Einzelfall, wie Lubinski (2004) resümierend feststellt:

In heterogenous collections of cognitive tests in a wide range of talent, general intelligence [das ist die allgemeine kognitive Fähigkeit; M.B.] accounts for roughly 50% of the common variance (quantitative, spatial, and verbal ability each account for approximately 8%–10% of the remaining common variance). (Lubinski, 2004, S. 98)

Es kann also festgehalten werden, dass in bisherigen ATI-Studien zur Wechselwirkung spezifischer Fähigkeiten und instruktionaler Maßnahmen die verwendeten manifesten Skalenscores hinsichtlich der Messung der spezifischen Fähigkeit nicht sehr reliabel waren (siehe hierzu auch Brunner & Süß, 2005). Jedoch wurde die allgemeine kognitive Fähigkeit (insbesondere bei Aggregation von Einzelskalen) deutlich reliabler gemessen (Lubinski, 2004, S. 99). Wenn nun keine Wechselwirkung mit spezifischen kognitiven Fähigkeiten gefunden wurde, kann dies wohl auch auf die Unreliabilität der zu Grunde liegenden Scores attribuiert werden.

Was könnte man nun tun, um ATI für den Mathematikunterricht zu untersuchen? Als Möglichkeit käme in Frage, die Wirkung spezifischer Formen des Mathematikunterrichts im Rahmen eines Strukturmodells zu untersuchen, in dem die kognitiven Fähigkeiten mathematischer Schülerleistung in Form eines Nested-Faktormodells konzeptionalisiert werden. Denn mit Spezifikation kognitiver Fähigkeiten als latente Variablen wird das Problem der Unreliabilität (aufgrund mangelnder interner Konsistenz) der manifesten Variablen hinsichtlich der spezifischen Fähigkeiten umgangen (DeShon, 1998).

Das Nested-Faktormodell erfüllt auch eine zweite Bedingung, die in neueren Arbeiten zum ATI-Forschungsprogramm aufgestellt wurde (Pellegrino, Baxter & Glaser, 1999; Snow & Swanson, 1992). Nach diesen Autoren ist es essenziell, dass die eingesetzten Assessmentinstrumente für die intendierten Lernergebnisse sensitiv sind. Beispielsweise zielt kognitiv herausfordernder Mathematikunterricht in erster Linie auf die Fähigkeit zum mathematischen Problemlösen ab. In der vorliegenden Arbeit wurde die mathematikspezifische Fähigkeit als die Fähigkeit zum mathematischen Problemlösen interpretiert. Folgt man dieser Interpretation, gehört die mathematikspezifische Fähigkeit zu den zentralen Ergebniskriterien, um (über generelle Effekte hinaus) Interaktionen zwischen Schülermerkmalen und kognitiv herausforderndem Mathematikunterricht zu untersuchen.

Im Zusammenhang mit möglichen ATI – aber natürlich auch generell – ist es eine sehr wichtige Aufgabe, das Leistungsniveau von Schülern präzise einzuschätzen. Mit dieser diagnostischen Frage beschäftigt sich der nächste Abschnitt.

### 10.3.3 Pädagogische Praxis

Der Frage nach der Diagnose von Schülerleistungen wird angesichts der Befunde in dieser Arbeit anhand von drei Aspekten nachgegangen, die für die pädagogische Praxis relevant sind: mathematische Lernschwierigkeiten, Hochbegabung in Mathematik sowie der Nutzen standardisierter Leistungstests für Lehrer. Zentral ist hierbei der Vergleich der Konzeptionalisierung der kognitiven Fähigkeiten mathematischer Schülerleistung in Form des Nested-Faktormodells und des Standardmodells. Zunächst werden mathematikspezifische Lernschwierigkeiten betrachtet.

*Mathematikspezifische Lernschwierigkeiten.* Ungefähr 6 Prozent der Kinder haben mathematische Lernschwierigkeiten, wovon der größte Anteil (70%) Jungen sind (Geary, 1993, 1994).

Ein zentrales formales Merkmal zur Diagnose von Lernschwierigkeiten ist ein Missverhältnis zwischen tatsächlicher Leistung und Leistungserwartung (Zielinski, 1996). Zum Beispiel gehen Swanson und Beebe-Frankenberger (2004, S. 475) davon aus, dass Schüler gefährdet sind, mathematische Lernschwierigkeiten zu entwickeln, deren mathematische Leistung unter dem 25. Perzentil liegt und deren Leistung gleichzeitig über dem 15. Perzentil bei der allgemeinen kognitiven Fähigkeit liegt. Unter der Annahme, dass die erzielte Leistung von der allgemeinen kognitiven Fähigkeit abhängig ist, liegen diese Kinder also unter dem Wert, der aufgrund ihrer allgemeinen kognitiven Fähigkeit zu erwarten wäre.

Die Diagnose von mathematischen Lernschwierigkeiten als Diskrepanz von Leistung und Leistungserwartung korrespondiert mit der Konzeptualisierung kognitiver Fähigkeiten in Form eines Nested-Faktormodells (aber auch mit einem *Higher-Order*-Modell, Modell 6a, Abschnitt 7.2). Die mathematikspezifische Fähigkeit, die spezifische verbale Fähigkeit und die allgemeine kognitive Fähigkeit sind wechselseitig unabhängig voneinander. Somit ist es möglich, trotz normal ausgeprägter allgemeiner kognitiver Fähigkeit und normal ausgeprägter spezifischer verbaler Fähigkeit (z.B. indiziert durch den PISA-Lesetest) niedrige Leistungen in der mathematikspezifischen Fähigkeit zu erzielen. Bei einer Konzeptualisierung kognitiver Fähigkeiten in Form eines Standardmodells ist die Unterscheidung dieses Profils aus theoretischer Sicht schwieriger zu rechtfertigen. Prinzipiell ist im Standardmodell, in dem die generelle mathematische, die generelle verbale und die fluide Fähigkeit wechselseitig (und in dieser Arbeit sehr hoch) interkorrelieren, davon auszugehen, dass Schüler mit sehr geringen Leistungen in der fluiden Fähigkeit auch nur sehr niedrige Leistungen in der generellen mathematischen Fähigkeit erzielen.

Zur Messung mathematischer Leistung können Aufgaben aus Schülerleistungstests und zur Messung der Leistungserwartung standardisierte Intelligenztests verwendet werden. Unter formalen Gesichtspunkten eignen sich diese psychometrischen Verfahren daher, um mathematikspezifische Lernschwierigkeiten zu diagnostizieren. Mit Blick auf die Ergebnisse in dieser Arbeit sind jedoch bei diesem Vorgehen zwei Einschränkungen zu konstatieren.

Erstens scheint es schwierig zu sein, die spezifischen kognitiven Fähigkeiten im Rahmen eines Nested-Faktormodells reliabel mit Summenscores zu erfassen (siehe hierzu auch Brunner & Süß, 2005). So lag beispielsweise in der vorliegenden Arbeit die Reliabilität des Summenwerts der vier stoffgebietsspezifischen Subskalen zur Messung der mathematikspezifischen Fähigkeit in der Gesamtstichprobe bei  $\omega = .13$  (siehe zur Berechnungsmethode Brunner & Süß, 2005; McDonald, 1999). Im Vergleich hierzu konnte die generelle mathematische Fähigkeit deutlich reliabler gemessen werden ( $\omega = .78$ ). Wie oben ausgeführt wurde, ist das Standardmodell jedoch aus theoretischer Sicht weniger geeignet, mathematische Lernschwierigkeiten zu beschreiben.

Zweitens ist mit Geary (1993, 1994) zu bedenken, dass mathematische Lernschwierigkeiten sehr heterogen sind und verschiedene Subformen annehmen können. Diese Subformen unterscheiden sich unter anderem danach, ob Schwierigkeiten beim Abrufen und beim Einprägen mathematischer Fakten bestehen, ob der Gebrauch mathematischer Prozeduren (z.B. Zählstrategien) Schwierigkeiten bereitet, oder ob es schwer fällt, figural-räumliche Fähigkeiten einzusetzen, um numerische Informationen zu repräsentieren und zu interpretieren. Zur Diagnose dieser spezifischen kognitiven Defizite sind herkömmliche mathematische Schülerleistungstests nicht geeignet, sondern es müssen Testverfahren eingesetzt werden, die sensitiv für die jeweiligen kognitiven Prozesse und eingesetzten Lösungsstrategien sind (Geary, 1994). Mit stark kognitionspsychologisch ausgerichteten Maßen mathematischer Fähigkeiten ist es möglich, die Schwierigkeiten von Schülern mit mathematischen Lernschwierigkeiten differenziert zu erfassen. Angesichts der zunächst psycho-

metrischen Definition von Lernschwierigkeiten als Missverhältnis von Leistung und Leistungserwartung wird deutlich, dass der in Abschnitt 10.3.1 angeregte Brückenschlag zwischen Psychometrie, Wissenspsychologie und Kognitionspsychologie auch relevant für die pädagogische Praxis ist.

Bevor also einem Schüler mathematische Lernschwierigkeiten diagnostiziert werden, legen die beiden genannten Einschränkungen nahe, verschiedene Erhebungsverfahren in einem mehrstufigen diagnostischen Prozess zu Rate zu ziehen. Im ersten Schritt des Prozesses kann die Diskrepanz zwischen Leistung und Leistungserwartung, wie sie mit standardisierten Tests gemessen wird, eine Gefährdung für die Entwicklung von mathematischen Lernschwierigkeiten indizieren. Angesichts der im Vergleich zur mathematikspezifischen Fähigkeit höheren Reliabilität der generellen mathematischen Fähigkeit scheint (zumindest vor dem Hintergrund dieser Arbeit) eine Konzeptionalisierung in Form des Standardmodells besser geeignet zu sein als das Nested-Faktormodell. Mit Hinzunahme kognitionspsychologischer Messverfahren kann dann im zweiten Schritt das theoretische „Defizit“ des Standardmodells mehr als ausgewogen werden.

*Mathematische Hochbegabung.* Die Frage nach der Konzeptionalisierung der kognitiven Fähigkeiten mathematischer Schülerleistung ist auch relevant für die Diagnose mathematischer Hochbegabung. Ein formales Charakteristikum mathematisch Hochbegabter ist, dass sie Ergebnisse in standardisierten Leistungstests erzielen, mit denen sie zu den Besten 2 Prozent ihrer Jahrgangskohorte gehören (Rost, 2001).

Um die tatsächlich vorzufindenden Profilausprägungen mathematischer Hochbegabung abzubilden, erscheint eine Konzeptualisierung der kognitiven Fähigkeiten mathematischer Schülerleistung in Form des Nested-Faktormodells gut geeignet. So werden in der Studie *Study of Mathematically Precocious Youth's* (SMPY, siehe Lubinski, Webb, Morelock & Benbow, 2001, für eine knappe Studienbeschreibung) drei Gruppen von hoch begabten Jugendlichen unterschieden: (spezifisch) mathematisch Hochbegabte, generell Hochbegabte und spezifisch sprachlich Hochbegabte (Dark & Benbow, 1990, 1991). Im Nested-Faktormodell sind die mathematikspezifische Fähigkeit, die spezifische verbale Fähigkeit und die allgemeine kognitive Fähigkeit wechselseitig unabhängig voneinander. Somit ist es – wie oben ausgeführt wurde – ohne weiteres möglich, diese Begabungsprofile zu unterscheiden. Bei einer Konzeptualisierung kognitiver Fähigkeit in Form eines Standardmodells ist die Unterscheidung dieser drei Profile aufgrund der Interkorrelationen der jeweiligen Fähigkeiten untereinander schwerer nachzuvollziehen: Entsprechend den Modellannahmen ist davon auszugehen, dass mathematisch Hochbegabte auch gleichzeitig verbal hoch begabt sind.

Für eine reliable Diagnose von mathematischer Hochbegabung sowie der Hochbegabungsprofile stellt sich jedoch wieder das gleiche Problem wie bei der Diagnose der mathematischen Lernschwierigkeiten. Es gilt abzuwägen, ob eine Konzeption vorzuziehen ist, die aus theoretischer Sicht (das ist das Nested-Faktormodell) oder aus Sicht der höheren Reliabilität (das ist das Standardmodell) zu präferieren ist.

*Nutzen standardisierter Leistungstests für Lehrer.* Über die Diagnose von Lernschwierigkeiten oder Hochbegabung hinaus, können standardisierte Leistungstests für Lehrkräfte in ihrem Berufsalltag nützlich sein. Denn vielen Lehrkräften fällt es schwer, das mit standardisierten Tests ermittelte Leistungsniveau ihrer Klasse (im Vergleich zu anderen Klassen) akkurat zu beurteilen (Schrader & Helmke, 1987; Spinath, 2005). Analog hierzu sind deutliche Unterschiede in der Notengebung unterschiedlicher Schulen derselben Schulform festzustellen. Für die gleiche Leistung bekommen

Schüler, die unterschiedliche Schulen derselben Schulform besuchen, unterschiedliche Noten (Baumert u.a., 2003; Lüdtke, Köller, Artelt, Stanat & Baumert, 2002).

Was folgt daraus? Die Ergebnisse aus Schülerleistungstests geben Lehrkräften einen aufschlussreichen Einblick in den gegenwärtigen Leistungsstand ihrer Schüler, den sie in zweierlei Hinsicht nutzen können:

Erstens haben Lehrkräfte durch die Rückmeldung des Leistungsstands ihrer Schüler die Möglichkeit, ihre Notengebung mit den Testergebnissen zu vergleichen und gegebenenfalls ihre Notengebung an die Testergebnisse anzugleichen. Dies wäre eine Möglichkeit – zusätzlich zu einer verstärkten Berücksichtigung pädagogisch-psychologischer Diagnostik in der Lehrerbildung, wie sie namhafte Forscher einfordern (Leutner, 2001; Schrader & Helmke, 2001; Tent, 2001) – die Leistungsdiagnose von Schülern zu verbessern.

Zweitens können Lehrer ihren Unterricht an das Leistungsniveau der Klasse oder spezifische Stärken und Schwächen ihrer Klasse besser adaptieren, wenn sie über den Leistungsstand ihrer Klasse informiert sind. Je differenzierter die Tests Zielkriterien des Mathematikunterrichts erfassen (z.B. in Form mathematischer Stoffgebiete oder den Arten des mathematischen Arbeitens), desto besser sollte dieser Adaptationsprozess gelingen (siehe auch Schrader & Helmke, 2001). Letztlich könnte so eine der Basisdimensionen der Unterrichtsqualität (vgl. Abschnitt 5.3) optimiert werden.

Ein zentraler Gesichtspunkt bei beiden Einsatzmöglichkeiten ist jedoch auch wieder die Frage, ob man die jeweiligen kognitiven Fähigkeiten mathematischer Schülerleistung unabhängig voneinander messen möchte (hier ist eine Konzeption in Form des Nested-Faktormodells zu präferieren) oder ob man an einem Fähigkeitskomplex (in Form des Standardmodells) interessiert ist. Letzteres hat den Vorteil der größeren Reliabilität, jedoch den Nachteil der „Vermischung“ mehrerer kognitiver Fähigkeiten.

#### 10.3.4 Schluss

Insgesamt gesehen haben die bisherigen Überlegungen einerseits deutlich gemacht, dass die Konzeptionalisierung der kognitiven Fähigkeiten mathematischer Schülerleistung in Form des Nested-Faktormodells aus theoretischer Sicht für viele Forschungsfragen interessant und auch zur Beschreibung von Problemen der pädagogischen Praxis relevant ist. Andererseits wurde auch klar, dass in der pädagogischen Praxis diese Konzeptionalisierung aufgrund von Reliabilitätsproblemen an Grenzen stoßen kann. So gesehen empfiehlt sich das Nested-Faktormodell vor allem für die Beantwortung von Forschungsfragen. Eine Konzeptionalisierung in Form des Standardmodells hat sich – wie auch die Ergebnisse dieser Arbeit zeigten – im Forschungskontext bewährt, es scheint aber am besten für eine reliable Einzelfalldiagnose geeignet zu sein.

Unabhängig von der gewählten Konzeptionalisierung belegen die Ergebnisse der vorliegenden Arbeit und die vorangegangenen Überlegungen die Tragfähigkeit des *coordinate measurement* (Messick, 1984) von Maßen mathematischer Schülerleistung und Indikatoren kognitiver Fähigkeiten im Forschungskontext und in der pädagogischen Praxis. Mit Richard Mayer kann daher abschließend festgestellt werden: „the study of intelligence and education provides an example of the fruitful interaction between the practical demands of educators and the basic research focus of cognitive scientists“ (Mayer, 2000, S. 519).

## Anhang

### Testdesign bei der PISA-2000-Studie

Tabelle 40: PISA-O: Testdesign für den internationalen Vergleich der PISA-2000-Studie

Minuten	Testhefte (internationaler Testteil [I], erster Testtag)								
	I1	I2	I3	I4	I5	I6	I7	I8	I9
15	R1	R2	R3	R4	R5	R6	R7	M4	N4
15	R1	R2	R3	R4	R5	R6	R7	M2	N2
15	R2	R3	R4	R5	R6	R7	R1	N1	M1
15	R2	R3	R4	R5	R6	R7	R1	N3	M3
15	<i>Pause</i>								
15	R4	R5	R6	R7	R1	R2	R3	R8	R9
15	R4	R5	R6	R7	R1	R2	R3	R8	R9
15	M1	N1	M3	N3	M2	N2	R8	R9	R8
15	M2	N2	M4	N4	M3	N3	R8	R9	R8
10	<i>Pause</i>								
06	Lesegeschwindigkeit								
30	Internationaler Schülerfragebogen								
10	Selbstreguliertes Lernen								
05	Computererfahrung								

R = Blöcke mit Leseaufgaben, M = Blöcke mit Mathematikaufgaben, N = Blöcke mit Naturwissenschaftsaufgaben.

Tabelle 41: PISA-E: Testdesign für den nationalen Bundesländervergleich der PISA-2000-Studie

Minuten	Testhefte (nationaler Testteil [N], zweiter Testtag)								
	N1	N2	N3	N4	N5	N6	N7	N8	N9
03	Testbezogene Motivation								
25	MA	MB	MC	MD	ME	MF	NA	NB	NC
25	NA	NB	NC	ND	NE	NF	MG	MH	MI
04	Testbezogene Motivation								
10	<i>Pause</i>								
20	Problemlösen								
30	Schlussfolgerndes Denken								
15	Gruppenaufgaben								
10	<i>Pause</i>								
20	Kooperation und Kommunikation								
30	Nationaler Schülerfragebogen								

M = Blöcke mit Mathematikaufgaben, N = Blöcke mit Naturwissenschaftsaufgaben.

## Stichprobenvergleich: PISA 2000 versus Stichprobe dieser Arbeit

Ist die Analysestichprobe dieser Arbeit noch mit der ursprünglichen PISA-Stichprobe von Neuntklässlern vergleichbar? Hierzu wurde die Stichprobe der Neuntklässler aus PISA 2000 mit der Analysestichprobe hinsichtlich mehrerer soziodemografischer, motivationaler und kognitiver Variablen verglichen (Tab. 42). Die entscheidende Frage bei den Analysen war, welche Verzerrung (relativer Bias) durch die Selektion entsteht. Als Maß für den relativen Bias wurde die Differenz zwischen dem ursprünglichen Mittelwert in der PISA-Stichprobe und dem Mittelwert in der Analysestichprobe durch den Mittelwert in der PISA-Stichprobe geteilt. Im Kontext von Simulationsstudien wird ein relativer Bias kleiner 5 Prozent als trivial, kleiner 10 Prozent als moderat und größer 10 Prozent als substantiell betrachtet (Flora & Curran, 2004). Weiterhin war von Interesse, ob das 95%-Konfidenzintervall um den Mittelwert in der Analysestichprobe den ursprünglichen Mittelwert enthielt. Die Rationale hierfür ist, dass der Mittelwert, der auf Grundlage der ursprünglichen Stichprobe geschätzt wurde, ein Punktschätzer für den Populationsmittelwert ist.<sup>31</sup> Enthält das 95%-Konfidenzintervall diesen Wert, dann kann man mit einer Irrtumswahrscheinlichkeit von 5 Prozent schlussfolgern, dass die Analysestichprobe aus einer Population mit einem solchen Mittelwert stammt.

Nicht alle Neuntklässler in der PISA-2000-Stichprobe hatten gültige Werte bei diesen Variablen (der Anteil fehlender Werte lag zwischen 91% für das Interesse am Lesen und 100% für die Schulformzugehörigkeit). Daher wurden fehlende Werte für alle analysierten Variablen zunächst mit dem Programm Norm (Schafer, 2000) imputiert, damit die Vergleiche auf derselben Stichprobe basieren. Im Imputationsmodell waren alle in Tabelle 42 aufgelisteten Variablen enthalten. Um mögliche Mittelwertunterschiede zwischen den Stichproben zu berücksichtigen, war auch eine Variable im Imputationsmodell enthalten, die indizierte, ob die Daten eines Schülers in der Analysestichprobe enthalten waren (codiert als 1) oder nicht (codiert als 0). Die Standardfehler zur Bestimmung der 95%-Konfidenzintervalle wurden mit Mplus 3.01 (Muthén & Muthén, 1998–2004b) mittels des Moduls *complex* berechnet. So wurde berücksichtigt, dass durch die mehrstufige Stichprobenziehung bei PISA eine genestete Datenstruktur entstand (Schüler sind genestet in Schulen) (zu dieser Problematik siehe z.B. Snijders & Bosker, 1999).

Insgesamt gesehen kann der relative Bias als trivial bezeichnet werden. Die größte Verzerrung resultierte für den Gesamtscore schlussfolgerndes Denken mit einer Unterschätzung des ursprünglichen Mittelwerts von 5 Prozent. Alle 95%-Konfidenzintervalle schließen die Mittelwerte der PISA-2000-Stichprobe mit ein. Diese Ergebnisse zeigen, dass die Analysestichprobe weitestgehend vergleichbar mit der ursprünglichen PISA-2000-Stichprobe der Neuntklässler war.

---

<sup>31</sup> Hierbei ist es wichtig zu beachten, dass dieser so geschätzte Populationsmittelwert gewissermaßen fiktiv ist, da er auf Grundlage ungewichteter Schülerdaten beruht. Dieser Wert entspricht also *nicht* dem tatsächlichen Populationsmittelwert für die Population der deutschen Neuntklässler aus dem Jahr 2000!

Tabelle 42: Analysen zur Vergleichbarkeit der Analysestichprobe mit der PISA-2000-Stichprobe

Vergleichsvariable	PISA	Analysestichprobe		Relativer Bias
	M	M [95%-KI]		
Alter	15,84	15,83	[15,82, 15,84]	0,00
Geschlecht	0,50	0,50	[0,49, 0,50]	0,01
ISEI	43,28	43,27	[42,84, 43,71]	0,00
Anzahl an Bücher	3,56	3,57	[3,5, 3,61]	0,00
Relativer Anteil Hauptschüler	0,16	0,15	[0,13, 0,17]	0,04
Relativer Anteil Realschüler	0,24	0,24	[0,22, 0,26]	0,00
Relativer Anteil Gymnasiasten	0,30	0,30	[0,28, 0,33]	-0,02
Relativer Anteil Gesamtschüler	0,13	0,13	[0,11, 0,15]	0,00
Relativer Anteil Schüler aus Schulen mit mehreren Bildungsgängen	0,17	0,18	[0,16, 0,20]	-0,01
Mathematik (national)	457,46	458,20	[454,85, 461,54]	0,00
Lesen	488,74	490,78	[487,07, 494,48]	0,00
Schlussfolgerndes Denken (Gesamtscore)	0,11	0,12	[0,09, 0,15]	-0,05
Interesse Lesen	2,50	2,50	[2,48, 2,52]	0,00
Interesse Mathematik	2,65	2,61	[2,58, 2,63]	0,02
Selbstkonzept Lesen	2,85	2,85	[2,84, 2,86]	0,00
Selbstkonzept Mathematik	2,78	2,74	[2,71, 2,76]	0,02

KI = Konfidenzintervall, ISEI = *International Socio-Economic Index of Occupational Status*.



## Testheftspezifische Korrekturparameter der WLE-Scores

Da die Testhefte zufällig den Schülern zum Bearbeiten vorgelegt wurden, sollten sich die mittleren Leistungen von Schülern, die ein bestimmtes Testheft bearbeiteten, *nicht* von den mittleren Leistungen von Schülern unterscheiden, die ein anderes Testheft bearbeiteten. Ist dies jedoch der Fall spricht man von *Testhefteffekten* (*booklet effects*, Adams & Wu, 2002, S. 157). Bei den nationalen Testheften wurden bei PISA 2000 keine Testhefteffekte dokumentiert, allerdings bei den internationalen Testheften (Adams & Wu, 2002): Als Erklärung boten Adams und Wu Effekte der Bearbeitungsreihenfolge an.

Aufgrund der Testhefteffekte wurden in der vorliegenden Arbeit die geschätzten WLE-Scores nachträglich für unterschiedliche mittlere Schwierigkeiten der Stoffgebiete in Abhängigkeit des bearbeiteten internationalen Testhefts korrigiert. Die mittleren Schwierigkeiten wurden mit dem Verfahren, wie es Adams und Wu beschreiben (2002, S. 160–161) getrennt für die mathematischen Stoffgebiete ermittelt. Die Testhefteffekte sind in Tabelle 43 dokumentiert.

Tabelle 43: Korrekturparameter der WLE-Scores in Abhängigkeit der internationalen Testhefte

Internationales Testheft	Algebra	Arithmetik	Geometrie	Stochastik
1	.05	.02	.06	.32
2	-.02	-.01	-.01	-.01
3	-.04	-.03	-.05	-.19
4	.02	.02	.06	.01
5	.01	.01	.02	.07
6	.05	-.03	-.01	-.02
7	.00	.00	.01	-.05
8	-.09	-.01	-.15	-.25
9	.01	.05	.07	.12

Allerdings sind hierbei mehrere Dinge zu bedenken (siehe Tab. 44):

Erstens, gemessen an der Varianz ( $\eta^2$ ), die durch mittlere Unterschiede zwischen den internationalen Testheften entstand, sind diese Unterschiede als eher marginal zu beurteilen. Zur Berechnung dieses Varianzanteils wurde eine einfaktorielle Varianzanalyse (mit neun Faktorstufen) durchgeführt, bei der das bearbeitete internationale Testheft die unabhängige Variable war.

Zweitens, trotz der Korrektur der WLE-Scores verschwanden die Mittelwertunterschiede zwischen den internationalen Testheften (insbesondere für Stochastik) nicht vollständig. Der Anteil der durch die Testhefte erklärten Varianz sank jedoch bei Stochastik etwas ab.

Drittens, traten auch unerwartet Mittelwertunterschiede zwischen den nationalen Testheften auf (siehe Tab. 44).

Viertens, dass Testhefteeffekte für die vorliegende Arbeit nur eine untergeordnete Rolle spielen: Die Arbeit fokussiert Varianzen, Kovarianzen und Mittelwerte für eine neu definierte Analysestichprobe der PISA-2000-Untersuchung und Schülersubgruppen aus dieser Stichprobe. Selbst wenn die WLE-Scores noch nicht vollständig für die Testheftschwierigkeiten korrigiert sind, betreffen mögliche Verzerrungen der WLE-Scores alle Schüler gleichermaßen, da die Testhefte zufällig zugeteilt wurden. Dies bedeutet schlimmstenfalls eine Verschlechterung der Reliabilität der WLE-Scores. Es hat aber auf das Verhältnis von Varianzen, Kovarianzen und Mittelwerte der untersuchten Subgruppen (z.B. Schulformen) keinen Einfluss.

Tabelle 44: Durch Testhefteeffekte erklärte Varianz ( $\eta^2$ ) der mathematischen Schülerleistung

Stoffgebiet	Internationale Testhefte		Nationale Testhefte	
	Unkorrigiert	Korrigiert	Unkorrigiert	Korrigiert
Algebra	.00	.00	.00	.00
Arithmetik	.00	.00	.00	.00
Geometrie	.00	.00	.00	.00
Stochastik	.06	.05	.02	.02

Vergleich von unkorrigierten WLE-Scores, mit WLE-Scores, die für Testhefteeffekte korrigiert wurden.

## Fehlende Werte bei den Maßen der kognitiven Fähigkeiten

In Tabelle 45 sind alle vorkommenden Muster gültiger und fehlender Werte bei den WLE-Scores zur Messung der kognitiven Fähigkeiten für die Gesamtstichprobe und für die jeweiligen Schulformen abgebildet (für eine nähere Beschreibung der WLE-Scores siehe Abschnitt 7.2.2). Die Tabelle ist so zu lesen, dass „x“ gültige und „0“ fehlende Werte indiziert. So konnte zum Beispiel für 29.231 Schüler der insgesamt 29.386 untersuchten Schüler das Muster gültiger und fehlender Werte mit „Muster 1“ beschrieben werden. Das heißt, diese Schüler hatten bei allen neun WLE-Scores gültige Werte.

Tabelle 45: Muster gültiger (x) und fehlender Werte (0) der WLE-Scores der kognitiven Fähigkeiten für die Gesamtstichprobe und in Abhängigkeit der Schulform

	Muster gültiger und fehlender Werte								
	1	2	3	4	5	6	7	8	9
Algebra	x	x	x	x	x	x	x	0	0
Arithmetik	x	x	x	x	x	x	x	0	0
Geometrie	x	x	x	x	x	x	x	0	0
Stochastik	x	x	x	x	x	x	x	0	0
Figurenanalogien	x	x	x	x	x	0	0	x	0
Wortanalogien	x	x	x	x	x	0	0	x	0
Information	x	x	x	x	0	x	x	x	x
Interpretation	x	x	0	0	0	x	0	x	x
Reflektion	x	0	x	0	0	x	0	x	x
Insgesamt	29.231	1	1	5	1	143	1	1	2
Hauptschulen	4.442			4	1	43			1
Realschulen	7.047	1				28			1
Gymnasien	8.877					31			
Gesamtschulen	3.692		1	1		23	1	1	
Schulen mit mehreren Bildungsgängen	5.173					18			

## Schulformspezifische deskriptive Statistiken

Tabelle 46: Deskriptive Statistiken getrennt für die Schulformen (SF) Hauptschule (HS), Realschule (RS) und Gymnasium (GY)

	SF	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. Algebra	HS	0,47								
	RS	0,60								
	GY	0,61								
2. Arithmetik	HS	0,22	0,63							
	RS	0,42	0,56							
	GY	0,50	0,58							
3. Geometrie	HS	0,22	0,41	0,47						
	RS	0,36	0,41	0,50						
	GY	0,42	0,42	0,52						
4. Stochastik	HS	0,18	0,19	0,17	0,50					
	RS	0,24	0,28	0,21	0,48					
	GY	0,32	0,33	0,26	0,53					
5. Figurenanalogien	HS	0,16	0,32	0,30	0,15	0,90				
	RS	0,32	0,31	0,34	0,18	0,90				
	GY	0,37	0,34	0,35	0,22	0,83				
6. Wortanalogien	HS	0,15	0,27	0,20	0,13	0,36	0,73			
	RS	0,31	0,32	0,30	0,18	0,42	0,80			
	GY	0,36	0,32	0,29	0,24	0,39	0,80			
7. Information	HS	0,16	0,34	0,30	0,24	0,29	0,29	0,49		
	RS	0,30	0,34	0,28	0,22	0,30	0,35	0,46		
	GY	0,30	0,28	0,25	0,22	0,27	0,34	0,15		
8. Interpretation	HS	0,16	0,35	0,30	0,26	0,29	0,32	0,61	0,65	
	RS	0,31	0,33	0,27	0,25	0,32	0,41	0,59	0,65	
	GY	0,30	0,29	0,24	0,23	0,28	0,38	0,49	0,44	
9. Reflektion	HS	0,15	0,29	0,22	0,19	0,23	0,25	0,46	0,57	0,17
	RS	0,23	0,25	0,20	0,17	0,24	0,31	0,43	0,54	0,44
	GY	0,21	0,19	0,17	0,15	0,19	0,26	0,34	0,42	0,29

noch Tabelle 46: Deskriptive Statistiken getrennt für die Schulformen (SF) Hauptschule (HS), Realschule (RS) und Gymnasium (GY)

	SF	1.	2.	3.	4.	5.	6.	7.	8.	9.
M	HS	-1,08	-0,84	-0,80	-0,32	-0,70	-0,87	-0,99	-0,99	-1,16
	RS	-0,12	-0,10	-0,13	0,19	0,00	-0,08	-0,10	-0,11	-0,22
	GY	1,28	0,91	0,82	0,79	0,70	0,95	0,69	0,75	0,67
SD	HS	0,93	1,21	1,32	1,19	0,97	1,00	0,98	0,90	1,09
	RS	1,13	1,17	1,29	1,07	0,96	1,04	0,92	0,84	1,02
	GY	1,20	1,23	1,20	1,09	0,89	1,14	0,86	0,82	0,91
#fehlend	HS	1	1	1	1	44	44	1	5	5
	RS	1	1	1	1	29	29	0	0	1
	GY	0	0	0	0	31	31	0	0	0

Korrelationen, Mittelwerte und Standardabweichungen beziehen sich auf die mit dem *Full Information Maximum Likelihood* (FIML) Schätzer ermittelten Stichprobenkennwerte. In der Diagonale sind die Reliabilitäten der WLE-Scores (kursiv dargestellt) eingetragen (die Berechnung erfolgte analog zu der in Abschnitt 7.2.2 beschriebenen Methode). Ein Grund für die teilweise sehr geringen Reliabilitäten der drei Indikatoren der verbalen Fähigkeiten (Information, Interpretation und Reflektion) ist möglicherweise, dass die Reliabilitäten durch das verwendete Berechnungsverfahren unterschätzt wurden, da die Fehlervarianzen überschätzt wurden (siehe hierzu Rost, 2004b). SF = Skalierungsfehler, M = Mittelwert, SD = Standardabweichung, #fehlend = absolute Anzahl fehlender Werte. Anzahl Hauptschüler/Realschüler/Gymnasiasten mit gültigen Werten bei allen Variablen 4.442/7.047/8.877.

## Modellparameter der Mehrgruppenfaktorenanalysen

Die unstandardisierten Modellparameter der manifesten Variablen sind für beide Modelle in Tabelle 47 dokumentiert. Bemerkenswert ist hierbei, dass gemessen an der unstandardisierten Faktorladung die mathematikspezifische Fähigkeit vornehmlich die Leistung in Algebra und weitaus geringer die anderen drei Stoffgebiete beeinflusste. Dieser Befund zeigte sich auch, wenn nicht der Algebra-WLE-Score als Referenzindikator gewählt wurde (und dessen Faktorladung auf 1,0 fixiert wurde).

Tabelle 47: Skalarinvariante Modellvarianten: Unstandardisierte ( $\lambda^*$ ) Faktorladungen, unstandardisierte Intercepts ( $\mu$ ) und schulformspezifische Residualvarianzen ( $\theta^2$ ) der manifesten Variablen

	Standardmodell							Nested-Faktormodell						
	$\lambda^*_{M'}$	$\lambda^*_{Gf}$	$\lambda^*_{V'}$	$\mu$	$\theta^2_{HS}$	$\theta^2_{RS}$	$\theta^2_{GY}$	$\lambda^*_{M'}$	$\lambda^*_g$	$\lambda^*_{V'}$	$\mu$	$\theta^2_{HS}$	$\theta^2_{RS}$	$\theta^2_{GY}$
Alg	1,00 <sup>a</sup>			-1,00	0,71	0,73	.67	1,00 <sup>a</sup>	1,00 <sup>a</sup>		-1,05	0,70	0,49	0,26
Ari	0,85			-0,92	1,06	0,81	.84	0,36	1,07		-0,91	1,04	0,87	0,93
Geo	0,78			-0,88	1,37	1,15	.93	0,26	1,03		-0,88	1,34	1,19	1,00
Sto	0,52			-0,33	1,26	0,97	.96	0,17	0,68		-0,33	1,26	0,99	0,98
FA		1,00 <sup>a</sup>		-0,80	0,61	0,55	.49		0,95		-0,67	0,66	0,60	0,51
WA		1,00 <sup>a</sup>		-0,76	0,69	0,64	.86		1,20		-0,84	0,71	0,63	0,79
Info			1,00 <sup>a</sup>	-0,99	0,46	0,41	.43		0,79	1,00 <sup>a</sup>	-0,98	0,48	0,42	0,44
Int			1,05	-1,01	0,24	0,23	.31		0,77	1,20	-1,01	0,23	0,22	0,31
Ref			1,03	-1,10	0,67	0,62	.57		0,74	1,22	-1,10	0,66	0,61	0,56

Alle Faktorladungen waren statistisch signifikant von Null verschieden. M = generelle mathematische Fähigkeit, Gf = fluide Fähigkeit, V = verbale Fähigkeit, HS = Hauptschule, RS = Realschule, GY = Gymnasium, M' = mathematikspezifische Fähigkeit, g = allgemeine kognitive Fähigkeit, V' = spezifische verbale Fähigkeit, Alg = Algebra, Ari = Arithmetik, Geo = Geometrie, Sto = Stochastik, FA = Figurenanalogien, WA = Wortanalogien, Info = Informationen ermitteln, Int = textbezogenes Interpretieren, Ref = Reflektieren und Bewerten.

<sup>a</sup> Die unstandardisierte Faktorladung dieser manifesten Variable wurde für alle Schulformen auf 1,00 fixiert.

In Tabelle 48 sind die schulformspezifischen Varianzen und Korrelationen der latenten Variablen eingetragen, die (mit Ausnahme der Varianz von M') nicht in Abschnitt 8.1 berichtet wurden.

Tabelle 48: Skalarinvariante Modellvarianten: Schulformspezifische Statistiken der latenten Variablen

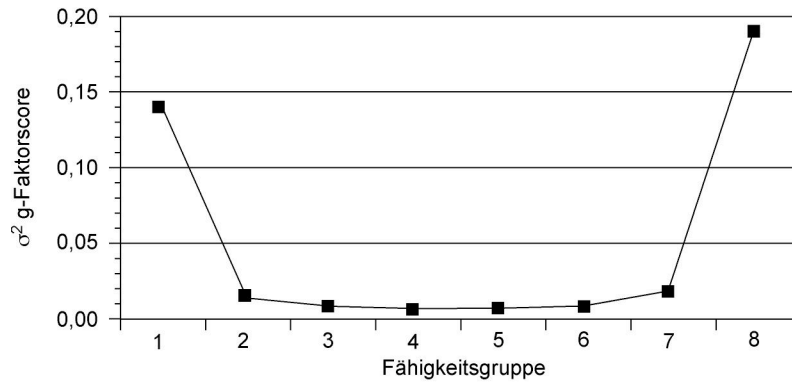
Schulform	$\sigma^2(M)/\sigma^2(M')$	$\sigma^2(Gf)/\sigma^2(g)$	$\sigma^2(V)/\sigma^2(V')$	$r(Gf, M)/r(g, M')$	$r(Gf, V)/r(g, V')$	$r(V, M)/r(V', M')$
<i>Standardmodell</i>						
Hauptschule	.38	.34	0.51	.75	.63	.66
Realschule	.68	.42	0.43	.77	.68	.63
Gymnasium	.84	.36	0.30	.83	.71	.58
<i>Nested-Faktormodell</i>						
Hauptschule	.05	.28	0.24	0 <sup>a</sup>	0 <sup>a</sup>	0 <sup>a</sup>
Realschule	.46	.36	0.17	0 <sup>a</sup>	0 <sup>a</sup>	0 <sup>a</sup>
Gymnasium	.75	.35	0.11	0 <sup>a</sup>	0 <sup>a</sup>	0 <sup>a</sup>

$\sigma^2$  = Varianz,  $r$  = Korrelation. Es ist zu beachten, dass die unstandardisierten Parameter nicht zwischen dem Standardmodell und dem Nested-Faktormodell verglichen werden können.

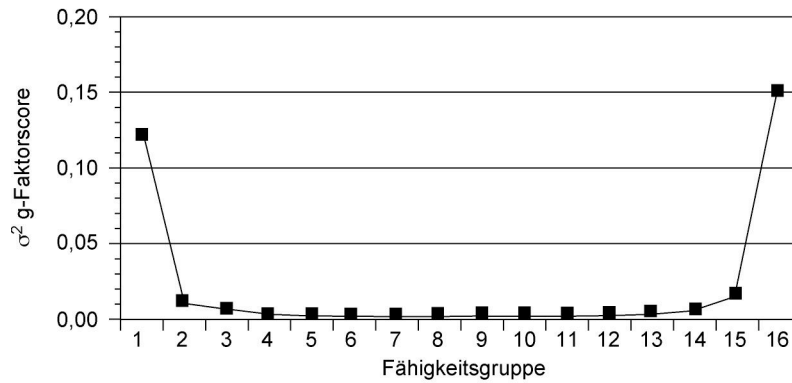
<sup>a</sup> Diese Modellparameter wurden auf Null fixiert.

## Fähigkeitsgruppenspezifische Varianz der Faktorscores der allgemeinen kognitiven Fähigkeit

8 Fähigkeitsgruppen



16 Fähigkeitsgruppen



32 Fähigkeitsgruppen

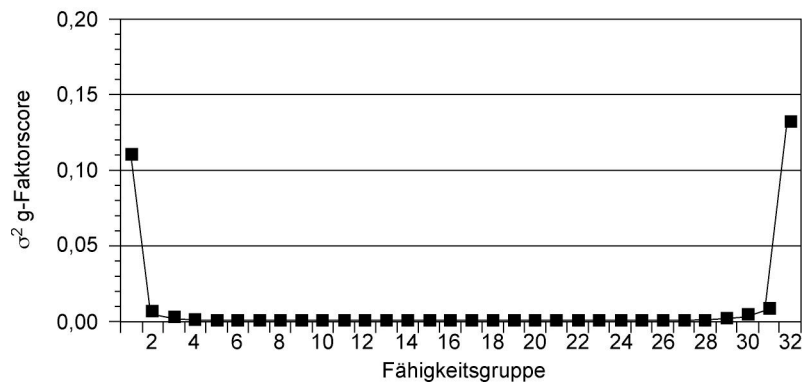
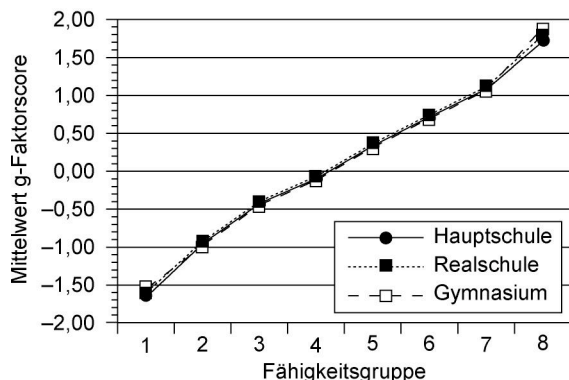


Abbildung 32: Varianzen der Faktorscores für die allgemeine kognitive Fähigkeit in Abhängigkeit des Fähigkeitsniveaus (in der Analytestichprobe)



(a) Mittelwerte



(b) Varianzen der Faktorscores

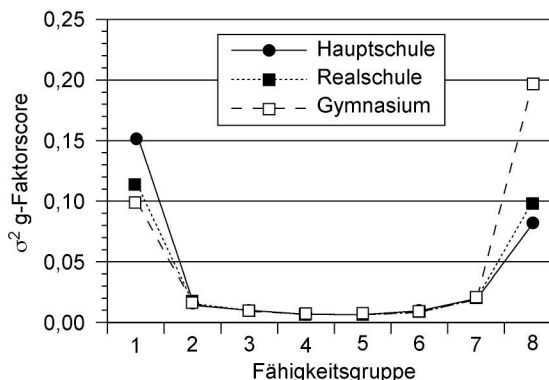


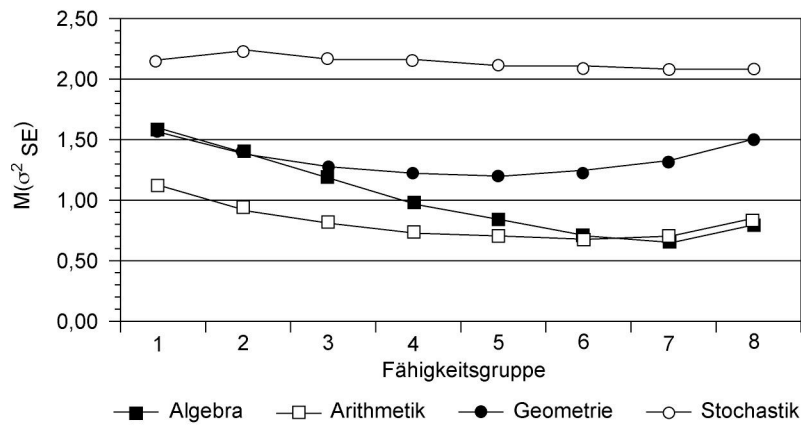
Abbildung 33: Mittelwerte und Varianzen der Faktorscores für allgemeine kognitive Fähigkeit in Abhängigkeit des Fähigkeitsniveaus und der Schulform

## Differenzierungshypothesen: Prüfung psychometrischer Gesichtspunkte

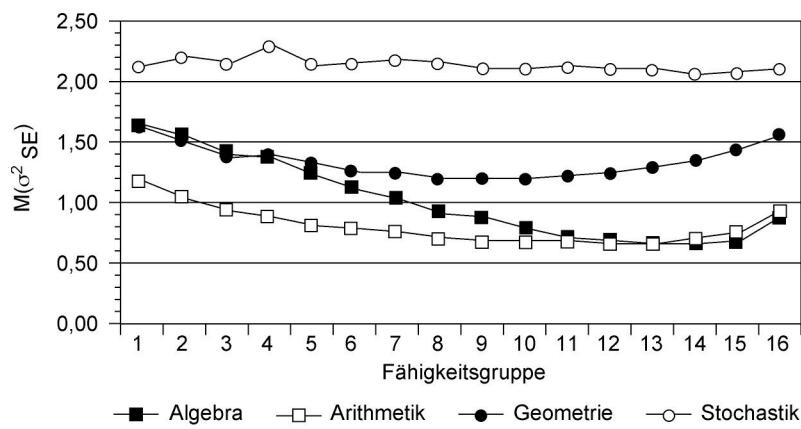
Die Frage nach Unterschieden der Heterogenität der Schüler hinsichtlich der mathematikspezifischen Fähigkeit wurde in dieser Arbeit anhand schulformspezifischer Unterschiede der Faktorvarianzen der mathematikspezifischen Fähigkeit beantwortet. Diese Faktorvarianzen basierten auf den Kovarianzen stoffgebietsspezifischer Maße mathematischer Schülerleistung. Ein fairer Vergleich der Kovariation zwischen Gruppen setzt voraus, dass die Maße mathematischer Schülerleistung ausreichend hohe schulformspezifische und fähigkeitsgruppenspezifische Reliabilitäten und Varianzen aufweisen. Waren diese Kriterien in der vorliegenden Arbeit erfüllt? Zur Beantwortung dieser Frage werden zunächst die Reliabilitäten betrachtet.

Als Maß der Reliabilität wurden die mittlere Schätzfehlervarianz der WLE-Scores für die Fähigkeitsgruppen und die schulformspezifischen Fähigkeitsgruppen berechnet (kleinere Varianzen indizieren eine höhere Reliabilität). Aus den Abbildungen 34 und 35 geht hervor, dass die mittlere Schätzfehlervarianz aller mathematischen WLE-Scores (mit Ausnahme von Algebra) in den Fähigkeitsgruppen wie auch in den schulformspezifischen Fähigkeitsgruppen weitestgehend konstant war.

### 8 Fähigkeitsgruppen



### 16 Fähigkeitsgruppen



### 32 Fähigkeitsgruppen

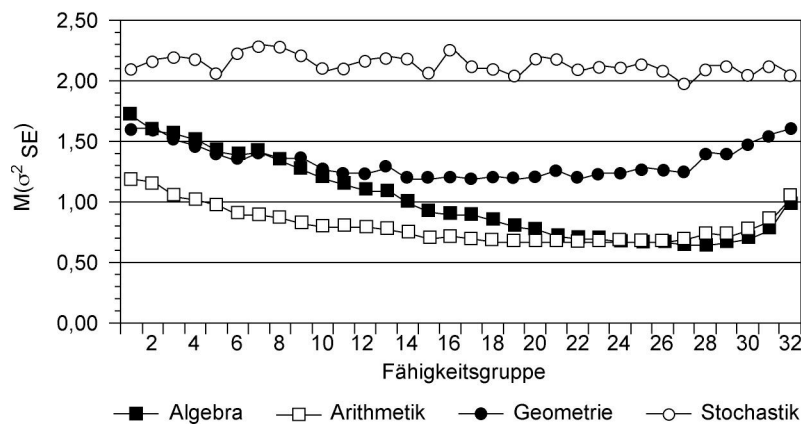


Abbildung 34: Mittlere Schätzfehlervarianz der stoffgebietspezifischen WLE-Scores in Abhängigkeit des Fähigkeitsniveaus (in der Analytestichprobe)

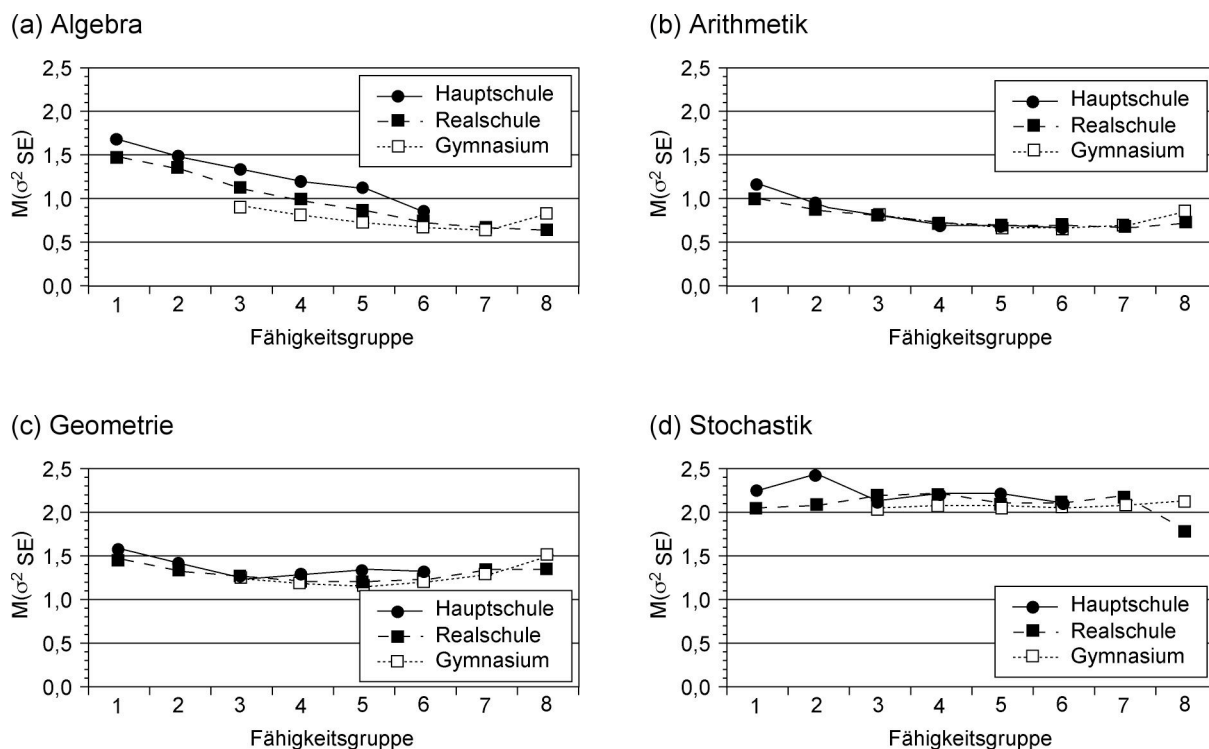


Abbildung 35: Mittlere Schätzfehlervarianz der stoffgebietspezifischen WLE-Scores in Abhängigkeit des schulformspezifischen Fähigkeitsniveaus

Angesichts dieser Ergebnisse stellt sich die Frage, ob die geringere Reliabilität des Algebra-WLE-Scores *generell* zu den beobachteten Unterschieden in der Differenzierung der mathematikspezifischen Fähigkeit führte. Gegen diese Annahme sprachen jedoch einige Gründe:

- Bei den Analysen zur generellen Differenzierungshypothese war auch ein Anstieg der Faktorvarianz von  $M'$  zwischen den Fähigkeitsgruppen 6 und 7 festzustellen, obwohl die mittlere Schätzfehlervarianz des Algebra-WLE-Scores bei diesen Fähigkeitsgruppen konstant blieb. Dies sprach gegen einen generellen Effekt der Reliabilität des Algebra-WLE-Scores.
- Die Analysen zur schulformspezifischen Differenzierung zeigten in der Fähigkeitsgruppe 6 eine Umkehrung der erwarteten Schulformrangfolge. Die mittlere Schätzfehlervarianz war jedoch bei Gymnasiasten geringer als bei Realschülern. Bei Hauptschülern war sie am größten. Dies widerspricht einem generellen Effekt der Reliabilität des Algebra-WLE-Scores.
- Die Reliabilität der Scores ist selbst Bestandteil der Differenzierung, denn die „wahre“ Varianz eines Scores setzt sich im Nested-Faktormodell aus drei Anteilen zusammen (siehe Abb. 17): allgemeine kognitive Fähigkeit, stoffgebietspezifische Fähigkeiten und mathematikspezifische Fähigkeit. Eine höhere Reliabilität des Algebra-WLE-Scores kann demnach auch auf einen stärkeren Einfluss der mathematikspezifischen Fähigkeit attribuiert werden. Reliabilitätsunterschiede können also ein Begleitphänomen von schulformspezifischen Unterschieden des Einflusses kognitiver Fähigkeiten sein.

Hierbei ist anzumerken, dass schulform- und fähigkeitsgruppenübergreifende konstante Reliabilitäten der WLE-Scores natürlich wünschenswert gewesen wären. Dies kann man erreichen, indem man gleiche schulformspezifische Testfunktionen durch die Itemzusammenstellung anstrebt (z.B.

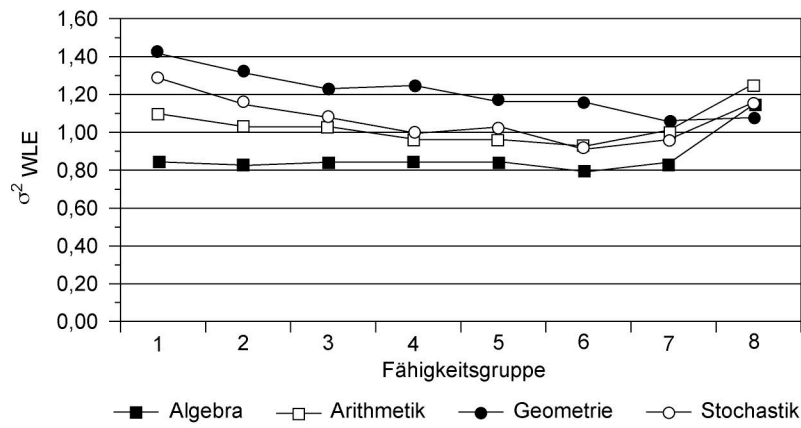
van der Linden, 2000). Hierbei ist jedoch zu bedenken, dass ein sehr großer Pool bearbeiteter Items notwendig ist, damit nicht auf Kosten der angestrebten konstanten Reliabilitäten, insgesamt gesehen zwar konstante, aber nur geringe Reliabilitäten der Scores erzielt werden. Weiterhin ist die inhaltliche Validität (American Educational Research Association, 1999) des neu gebildeten Tests gefährdet, wenn nicht alle mathematikdidaktischen Gesichtspunkte bei der Testzusammenstellung berücksichtigt werden. Aufgrund der begrenzten Gesamtanzahl von 117 Aufgaben im PISA-2000-Mathematiktest von denen Schüler maximal 41 bearbeiteten (vgl. Tab. 5), erschien es in der vorliegenden Arbeit nicht möglich die Anforderungen der inhaltlichen Validität gleichzeitig mit den schulformübergreifenden konstanten und ausreichend hohen Reliabilitäten für jedes Stoffgebiet zu vereinbaren.

Neben Reliabilitätsunterschieden kann eine potenzielle Varianzrestriktion für den Effekt schulformspezifischer Unterschiede des Einflusses verantwortlich sein. Wie aus den Abbildungen 36 und 37 hervorgeht, war dies aber unplausibel:

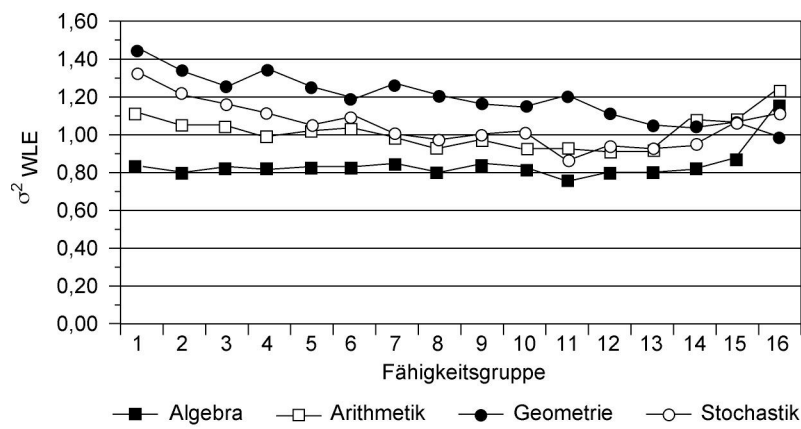
- Wie zum Beispiel Abbildung 36 für 8 Fähigkeitsgruppen zeigt, nahm die Varianz mit Ausnahme des Algebra-WLE-Scores bis zu Fähigkeitsgruppe 7 sogar tendenziell ab. Dies würde – unter der Annahme eines Effekts der Varianzrestriktion – dafür sprechen, dass die beobachtete generelle Differenzierung der mathematikspezifischen Fähigkeit eine untere Grenze für diese Differenzierung darstellt.
- Aus den Abbildungen 37a bis 37d, die die Varianzen in Abhängigkeit der schulformspezifischen Fähigkeitsgruppen zeigen, ist nicht erkennbar, dass die Varianz der mathematischen WLE-Scores eingeschränkt war. Die beobachteten Schulformunterschiede können also kaum auf eine Varianzrestriktion attribuiert werden.

Zusammenfassend kann man festhalten, dass weder Reliabilitätsunterschiede noch potenzielle Varianzeinschränkungen die generelle wie auch die schulformspezifische Differenzierung der mathematikspezifischen Fähigkeit generell erklären konnten.

### 8 Fähigkeitsgruppen



### 16 Fähigkeitsgruppen



### 32 Fähigkeitsgruppen

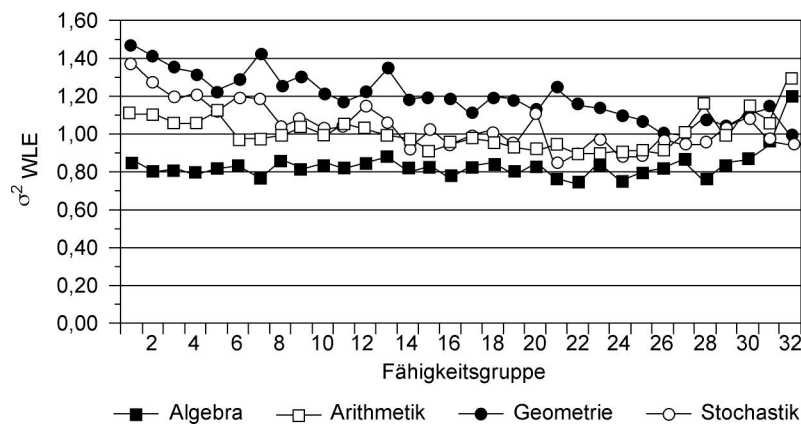
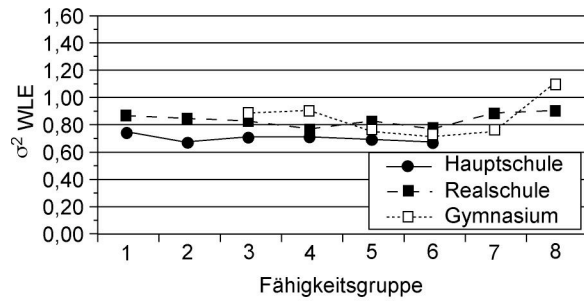
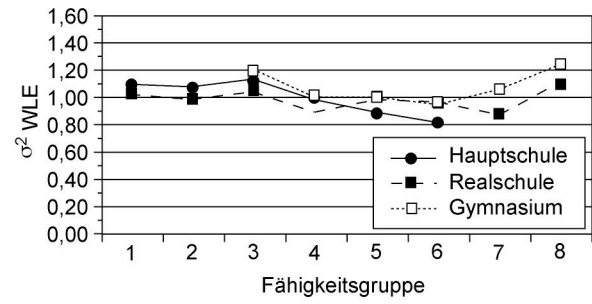


Abbildung 36: Varianzen der stoffgebietspezifischen WLE-Scores in Abhängigkeit des Fähigkeitsniveaus (in der Analytestichprobe)

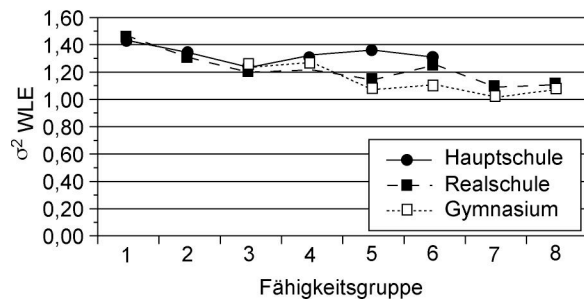
(a) Algebra



(b) Arithmetik



(c) Geometrie



(d) Stochastik

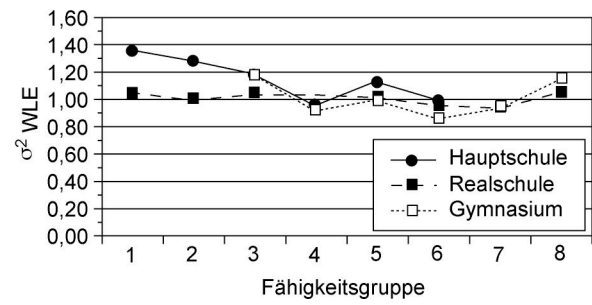


Abbildung 37: Varianzen der stoffgebietspezifischen WLE-Scores in Abhängigkeit des schulformspezifischen Fähigkeitsniveaus

## Interkorrelationen der Maße kognitiver Fähigkeiten bei Kontrolle der Schulformzugehörigkeit

Tabelle 49: Vergleich der deskriptiven Statistiken mit und ohne Kontrolle der Schulformzugehörigkeit (SF)

	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. Algebra		0,58	0,52	0,40	0,51	0,52	0,52	0,54	0,47
2. Arithmetik	0,42		0,53	0,39	0,48	0,48	0,50	0,52	0,44
3. Geometrie	0,36	0,42		0,33	0,47	0,44	0,44	0,45	0,38
4. Stochastik	0,26	0,28	0,23		0,32	0,33	0,37	0,39	0,32
5. Figurenanalogien	0,33	0,34	0,35	0,20		0,55	0,48	0,50	0,43
6. Wortanalogien	0,32	0,32	0,29	0,20	0,41		0,53	0,57	0,49
7. Information	0,30	0,34	0,29	0,25	0,32	0,36		0,71	0,59
8. Interpretation	0,31	0,35	0,29	0,26	0,34	0,40	0,59		0,67
9. Reflektion	0,23	0,26	0,22	0,18	0,25	0,30	0,44	0,53	
SD (ohne Kontrolle SF)	1,41	1,36	1,41	1,18	1,08	1,27	1,09	1,05	1,20
SD (mit Kontrolle SF)	1,13	1,20	1,28	1,11	0,96	1,09	1,02	0,93	0,87
Bias	0,80	0,88	0,91	0,94	0,89	0,86	0,94	0,89	0,72

Unterhalb der Diagonale sind die Korrelationen bei Kontrolle der Schulformzugehörigkeit eingetragen. Diese Statistiken basieren auf einer EM-Schätzung (siehe Abschnitt 9.1.2). Über der Diagonale stehen die Korrelationen ohne Kontrolle der Schulformzugehörigkeit. Diese basieren auf den mit dem *Full Information Maximum Likelihood* (FIML) Schätzer ermittelten Stichprobenkennwerten (siehe Abschnitt 7.2). SD = Standardabweichung, Bias = SD (mit Kontrolle SF) geteilt durch SD (ohne Kontrolle SF).

## Vergleich der Korrelationen zwischen Schülermerkmalen und kognitiven Fähigkeiten in Abhängigkeit verschiedener faktorieller Strukturmodelle

Tabelle 50: Vergleich der Korrelationen zwischen Schülermerkmalen und den kognitiven Fähigkeiten mathematischer Schülerleistung in Abhängigkeit verschiedener faktorieller Strukturmodelle

Schülermerkmal	Kognitive Fähigkeiten								
	Gf	g	g <sub>HO</sub>	M	M'	M' <sub>HO</sub>	V	V'	V' <sub>HO</sub>
Geschlecht	-.04	-.05	-.05	.13	.42	.17	-.12	-.17	-.08
ISEI	.33	.35	.34	.34	.05	.03	.35	.09	.05
Anzahl an Bücher	.40	.41	.41	.38	.01	.00	.41	.11	.05
Deutschnote	-.15	-.16	-.16	-.10	.10	.04	-.20	-.13	-.06
Mathematiknote	-.23	-.23	-.24	-.31	-.24	-.08	-.13	.14	.08
Interesse am Lesen	.29	.31	.31	.20	-.21	-.09	.36	.20	.10
Mathematisches Interesse	.10	.09	.11	.22	.34	.12	-.05	-.26	-.15
Verbales Selbstkonzept	.13	.15	.14	.07	-.17	-.06	.25	.25	.14
Mathematisches Selbstkonzept	.22	.21	.23	.34	.36	.12	.06	-.23	-.14

Fähigkeiten im Standardmodell: Gf = fluide Fähigkeit, M = generelle mathematische Fähigkeit, V = generelle verbale Fähigkeit. Fähigkeiten im Nested-Faktormodell: g = allgemeine kognitive Fähigkeit, M' = mathematikspezifische Fähigkeit, V' = spezifische verbale Fähigkeit. Fähigkeiten im *Higher-Order*-Modell: g<sub>HO</sub> = allgemeine kognitive Fähigkeit, M'<sub>HO</sub> = residualisierte mathematische Fähigkeit, V'<sub>HO</sub> = residualisierte verbale Fähigkeit. Geschlecht (Mädchen = 0, Jungen = 1), ISEI = *International Socio-Economic Index of Occupational Status*.



## Interkorrelationen der untersuchten Schülervariablen

Tabelle 51: Interkorrelationen der kognitiven Fähigkeiten, der Schulformzugehörigkeit und der untersuchten Schülervariablen für das Standardmodell und das Nested-Faktormodell

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.
1. M/M'		.00 <sup>a</sup>	.00 <sup>a</sup>	-.22	.35	.37	-.18	.11	-.25	.43	.01	.05	-.05	.12	-.14	.02
2. Gf/g	.89		.00 <sup>a</sup>	.31	.09	.21	.15	-.16	-.23	-.05	.41	.35	-.06	.63	-.12	-.19
3. V/V'	.80	.83		.20	-.26	-.23	.25	-.13	.14	-.17	.12	.09	.04	.08	.00	-.01
4. INT-L	.19	.29	.36		-.05	-.05	.35	-.21	.00	-.30	.30	.14	-.07	.22	-.02	-.09
5. INT-M	.22	.10	-.05	-.05		.87	-.15	.04	-.47	.23	-.01	-.02	-.01	-.05	.00	.01
6. SK-M	.33	.22	.06	-.04	.87		-.13	-.02	-.62	.22	.03	.03	-.02	.00	.00	.00
7. SK-V	.07	.13	.25	.35	-.15	-.13		-.55	.03	-.25	.16	.09	-.06	.13	.00	-.02
8. D-Note	-.10	-.15	-.20	-.21	.04	-.02	-.54		.31	.20	-.12	-.07	.00 <sup>b</sup>	.00 <sup>b</sup>	.00 <sup>b</sup>	.00 <sup>b</sup>
9. M-Note	-.30	-.23	-.13	.00	-.47	-.62	.03	.31		-.09	-.05	-.06	.00 <sup>b</sup>	.00 <sup>b</sup>	.00 <sup>b</sup>	.00 <sup>b</sup>
10. Geschl.	.13	-.04	-.12	-.30	.23	.22	-.25	.20	-.08		-.04	.03	.01	-.07	.01	.02
11. Anz.-B.	.38	.39	.41	.30	-.01	.03	.16	-.12	-.05	-.04		.38	-.06	.34	-.03	-.14
12. ISEI	.34	.33	.35	.14	-.02	.03	.09	-.07	-.06	.03	.38		-.05	.34	-.02	-.16
13. DRS	-.08	-.05	-.03	-.07	-.01	-.02	-.06	.00 <sup>b</sup>	.00 <sup>b</sup>	.01	-.06	-.05		-.37	-.21	-.26
14. DGY	.64	.59	.58	.22	-.05	.00	.13	.00 <sup>b</sup>	.00 <sup>b</sup>	-.07	.34	.34	-.37		-.25	-.31
15. DGS	-.17	-.11	-.10	-.02	.00	.00	.00	.00 <sup>b</sup>	.00 <sup>b</sup>	.01	-.03	-.02	-.21	-.25		-.18
16. DMGB	-.16	-.19	-.17	-.09	.01	.00	-.02	.00 <sup>b</sup>	.00 <sup>b</sup>	.02	-.14	-.16	-.26	-.31	-.18	

Unter der Diagonale sind die Interkorrelationen für das Standardmodell eingetragen. Über der Diagonale stehen die Interkorrelationen für das Nested-Faktormodell. M = generelle mathematische Fähigkeit, M' = mathematikspezifische Fähigkeit, Gf = fluide Fähigkeit, g = allgemeine kognitive Fähigkeit, V = generelle verbale Fähigkeit, V' = spezifische verbale Fähigkeit, INT-L = Interesse am Lesen, INT-M = mathematisches Interesse, SK-M = mathematisches Selbstkonzept, SK-V = verbales Selbstkonzept, D-Note = Deutschnote, M-Note = Mathematiknote, Geschl. = Geschlecht (Mädchen = 0, Jungen = 1), Anz.-B. = Anzahl an Büchern. ISEI = *International Socio-Economic Index of Occupational Status*, DRS = Dummyvariable für die Zugehörigkeit zur Realschule (0 = andere Schulformen, 1 = Realschule), DGY = Dummyvariable für die Zugehörigkeit zum Gymnasium (0 = andere Schulformen, 1 = Gymnasium), DGS = Dummyvariable für die Zugehörigkeit zur Gesamtschule (0 = andere Schulformen, 1 = Gesamtschule), DMGB = Dummyvariable für die Zugehörigkeit zur Schule mit mehreren Bildungsgängen (0 = andere Schulformen, 1 = Schule mit mehreren Bildungsgängen).

<sup>a</sup> Diese Modellparameter wurden auf Null fixiert.

<sup>b</sup> Diese Korrelationen waren „Null“, weil die Noten innerhalb von Schulen zentriert wurden.

## Regressionsmodelle zur Vorhersage von kognitiven Fähigkeiten und Schulnoten

Die durchgeführten Analysen zu den Geschlechterdifferenzen, zum familiären Hintergrund und den Schulnoten konzentrierten sich auf die bivariaten Zusammenhänge zwischen den kognitiven Fähigkeiten mathematischer Schülerleistung und diesen Schülermerkmalen. Auch bei den Analysen zu den domänenspezifischen Selbstkonzepten und Interessen wurde nur eine begrenzte Anzahl an Prädiktoren gleichzeitig analysiert. Deshalb wurde ein explorativer Blick auf die multi-

variablen Zusammenhänge zwischen allen untersuchten Schülermerkmalen (inkl. der Schulformzugehörigkeit, vgl. Kap. 8) geworfen (siehe auch Tabelle 51 zu einer Interkorrelation aller untersuchten Schülervariablen inkl. der Schulformzugehörigkeit).

Hierzu wurden die Schulnoten und die kognitiven Fähigkeiten mathematischer Schülerleistung untersucht. Angesichts der Multikolarität der Variablen (siehe Tabelle 51), die eine weitergehende Interpretation der Regressionsgewichte weitestgehend unterbindet (Cohen & Cohen, 1983), ist vor allem interessant, wie viel Prozent der Varianz insgesamt erklärt werden konnte.

Bei den kognitiven Fähigkeiten lagen die Anteile erklärter Varianz zwischen 55 Prozent (Gf und V) und 64 Prozent (M), wohingegen sie im Nested-Faktormodell zwischen 33 Prozent (V') und 65 Prozent (g) lagen (Tab. 52).

Tabelle 52: Vorhersage der kognitiven Fähigkeiten mathematischer Schülerleistung durch die analysierten Schülermerkmale und die Schulformzugehörigkeit

Prädiktor	Standardmodell			Nested-Faktormodell		
	Abhängige Variable			Abhängige Variable		
	Gf	M	V	g	M'	V'
INT-L	.14	.07	.17	.14	-.11	.18
INT-M	-.19	-.13	-.26	-.21	.09	-.27
SK-V	-.03	-.02	.06	.00	-.08	.15
SK-M	.34	.35	.28	.35	.29	.06
D-Note	-.07	-.10	-.04	-.08	-.14	.05
M-Note	-.09	-.07	-.08	-.09	.00	-.05
Geschl.	.01	.15	.00	.06	.34	-.12
Anz.-B.	.12	.10	.11	.12	.05	.06
ISEI	.04	.05	.06	.06	.01	.05
DRS	.34	.32	.38	.38	.06	.25
DGY	.74	.80	.73	.83	.41	.31
DGS	.19	.15	.21	.20	-.06	.17
DMGB	.20	.22	.24	.24	.07	.15
R <sup>2</sup>	.55	.64	.55	.65	.59	.33

In beiden Modellen wurden die Faktorladungen der Maße der kognitiven Fähigkeiten auf die unstandardisierten Parameter fixiert, die bei der Analyse in Abschnitt 7.2 ermittelt wurden. Durch diese Invarianzspezifikation waren die jeweiligen Faktoren äquivalent zu den Analysen in Abschnitt 7.2. Gf = fluide Fähigkeit, M = generelle mathematische Fähigkeit, V = generelle verbale Fähigkeit, g = allgemeine kognitive Fähigkeit, M' = mathematikspezifische Fähigkeit, V' = spezifische verbale Fähigkeit, INT-L = Interesse am Lesen, INT-M = mathematisches Interesse, SK-V = verbales Selbstkonzept, SK-M = mathematisches Selbstkonzept, D-Note = Deutschnote, M-Note = Mathematiknote, Geschl. = Geschlecht (Mädchen = 0, Jungen = 1), Anz.-B. = Anzahl an Büchern, ISEI = *International Socio-Economic Index of Occupational Status*, DRS = Dummyvariable für die Zugehörigkeit zur Realschule (0 = andere Schulformen, 1 = Realschule), DGY = Dummyvariable für die Zugehörigkeit zum Gymnasium (0 = andere Schulformen, 1 = Gymnasium), DGS = Dummyvariable für die Zugehörigkeit zur Gesamtschule (0 = andere Schulformen, 1 = Gesamtschule), DMGB = Dummyvariable für die Zugehörigkeit zur Schule mit mehreren Bildungsgängen (0 = andere Schulformen, 1 = Schule mit mehreren Bildungsgängen).

Mit Blick auf die Schulnoten waren die Anteile aufgeklärter Varianz in beiden Modellen fast identisch (Tab. 53). Die Werte lagen hierbei im Standardmodell (Nested-Faktormodell) bei 34 Prozent (35%) bei der Deutschnote und 43 Prozent (43%) bei der Mathematiknote.

Tabelle 53: Vorhersage der Schulnoten durch die kognitiven Fähigkeiten mathematischer Schülerleistung, die analysierten Schülermerkmale und die Schulformzugehörigkeit

Prädiktor	Standardmodell		Prädiktor	Nested-Faktormodell	
	Abhängige Variable			Abhängige Variable	
	D-Note	M-Note		D-Note	M-Note
M	-.10	-.26	M'	-.02	-.09
V	-.03	.07	V'	-.01	.03
Gf	-.08	.00	g	-.19	-.17
INT-L	.02	.01	INT-L	.03	.01
INT-M	.13	.27	INT-M	.13	.27
SK-V	-.53	-.04	SK-V	-.53	-.04
SK-M	-.17	-.79	SK-M	-.17	-.80
Geschl.	.10	.07	Geschl.	.10	.07
Anz.-B.	-.02	.01	Anz.-B.	-.02	.01
ISEI	-.03	-.02	ISEI	-.03	-.02
DRS	.09	.05	DRS	.10	.05
DGY	.29	.18	DGY	.29	.17
DGS	.07	.02	DGS	.08	.02
DMBG	.07	.03	DMBG	.07	.03
R <sup>2</sup>	.34	.43	R <sup>2</sup>	.35	.43

In beiden Modellen wurden die Faktorladungen der Maße der kognitiven Fähigkeiten auf die unstandardisierten Parameter fixiert, die bei der Analyse in Abschnitt 7.2 ermittelt wurden. Durch diese Invarianzspezifikation waren die jeweiligen Faktoren äquivalent zu den Analysen in Abschnitt 7.2. D-Note = Deutschnote, M-Note = Mathematiknote, M = generelle mathematische Fähigkeit, M' = mathematikspezifische Fähigkeit, V = generelle verbale Fähigkeit, V' = spezifische verbale Fähigkeit, Gf = fluide Fähigkeit, g = allgemeine kognitive Fähigkeit, INT-L = Interesse am Lesen, INT-M = mathematisches Interesse, SK-V = verbales Selbstkonzept, SK-M = mathematisches Selbstkonzept, Geschl. = Geschlecht (Mädchen = 0, Jungen = 1), Anz.-B. = Anzahl an Büchern, ISEI = *International Socio-Economic Index of Occupational Status*, DRS = Dummyvariable für die Zugehörigkeit zur Realschule (0 = andere Schulformen, 1 = Realschule), DGY = Dummyvariable für die Zugehörigkeit zum Gymnasium (0 = andere Schulformen, 1 = Gymnasium), DGS = Dummyvariable für die Zugehörigkeit zur Gesamtschule (0 = andere Schulformen, 1 = Gesamtschule), DMBG = Dummyvariable für die Zugehörigkeit zur Schule mit mehreren Bildungsgängen (0 = andere Schulformen, 1 = Schule mit mehreren Bildungsgängen).

## Literatur

- Abad, F. J., Colom, R., Juan-Espinosa, M., & Garcia, L. F. (2003). Intelligence differentiation in adult samples. *Intelligence, 31*, 157–166.
- Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102*, 3–27.
- Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General, 117* (3), 288–318.
- Ackerman, P. L. (1989). Individual differences and skill acquisition. In P. L. Ackerman, R. J. Sternberg & R. Glaser (Eds.), *Learning and individual differences: Advances in theory and research* (pp. 165–217). New York: Freeman.
- Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence, 22*, 227–257.
- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General, 131* (4), 567–589.
- Ackerman, P. L., & Kanfer, R. (1993). Integrating laboratory and field study for improving selection: Development of a battery for predicting air traffic controller success. *Journal of Applied Psychology, 78*, 413–432.
- Ackerman, P. L., Kyllonen, P. C., & Roberts, R. D. (Eds.). (1999). *Learning and individual differences: Process, trait and content determinants*. Washington, DC: APA Books.
- Ackerman, P. L., Sternberg, R. J., & Glaser, R. (Eds.). (1989). *Learning and individual differences: Advances in theory and research*. New York: Freeman.
- Adams, R., & Wu, M. (2002). *PISA 2000 Technical Report*. Paris: OECD.
- Ahnen, D. (2004). Vorwort der Präsidentin der Kultusministerkonferenz. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (pp. 11–12). Münster: Waxmann.
- American Educational Research Association, A.P.A., National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.
- Anastasi, A. (1968). *Psychological testing* (3rd ed.). New York: Macmillan.
- Anastasi, A. (1970). On the formation of psychological traits. *American Psychologist, 25*, 899–910.
- Anastasi, A. (1981). Coaching, test sophistication, and developed abilities. *American Psychologist, 36* (10), 1086–1093.
- Anastasi, A. (1986). Experiential structuring of psychological traits. *Developmental Review, 6*, 181–202.
- Anderson, J. R. (1993). Problem solving and learning. *American Psychologist, 48*, 35–44.
- Anderson, J. R. (2001). *Kognitive Psychologie* (3rd ed.). Heidelberg: Spektrum Akademischer Verlag.
- Artelt, C., Brunner, M., Schneider, W., Prenzel, M., & Neubrand, M. (2003). Literacy oder Lehrplanvalidität? Ländervergleiche auf der Basis lehrplanoptimierter PISA-Tests. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (S. 77–108). Opladen: Leske + Budrich.
- Artelt, C., Stanat, P., Schneider, W., & Schiefele, U. (2001). Lesekompetenz: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Eds.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 69–137). Opladen: Leske + Budrich.
- Avenarius, H., Ditton, H., Döbert, H., Klemm, K., Klieme, E., Rürup, M., Tenorth, H.-E., Weishaupt, H., & Weiß, M. (2003). *Bildungsbericht für Deutschland. Erste Befunde*. Opladen: Leske + Budrich.
- Ball, D., Lubienski, S. T., & Mewborn, D. S. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 433–456). New York: Macmillan.
- Baltes, P. B., Lindenberger, U., & Staudinger, U. M. (1998). Life-span theory in developmental psychology. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (pp. 1029–1143). New York: Wiley.
- Baltes, P. B., Nesselroade, J. R., & Cornelius, S. W. (1978). Multivariate antecedents of structural change in development: A simulation of cumulative environmental patterns. *Multivariate Behavioral Research, 13*, 127–152.
- Baumert, J. (2001). Internationale Schulleistungsvergleiche. In D. H. Rost (Ed.), *Handwörterbuch Pädagogische Psychologie* (2. Aufl., S. 294–303). Weinheim: Psychologie Verlags Union.
- Baumert, J., & Artelt, C. (2002). Bereichsübergreifende Perspektiven. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Die Länder der Bundesrepublik im Vergleich* (S. 219–236). Opladen: Leske + Budrich.

- Baumert, J., & Artelt, C. (2003). Konzeption und technische Grundlagen der Studie. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (S. 11–50). Opladen: Leske + Budrich.
- Baumert, J., Artelt, C., Carstensen, C. H., Sibberns, H., & Stanat, P. (2002). Untersuchungsgegenstand, Fragestellungen und technische Grundlagen der Studie. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Die Länder der Bundesrepublik im Vergleich* (S. 11–38). Opladen: Leske + Budrich.
- Baumert, J., Bos, W., & Lehmann, R. (Hrsg.). (2000). *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn: Bd 2. Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit*. Opladen: Leske + Budrich.
- Baumert, J., & Köller, O. (2000). Unterrichtsgestaltung, verständnisvolles Lernen und multiple Zielerreichung im Mathematik- und Physikunterricht der gymnasialen Oberstufe. In J. Baumert, W. Bos & R. Lehmann (Hrsg.), *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn: Bd 2. Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe* (S. 271–315). Opladen: Leske + Budrich.
- Baumert, J., Köller, O., Lehrke, M., & Brockmann, J. (2000). Anlage und Durchführung der Dritten Internationalen Mathematik- und Naturwissenschaftsstudie zur Sekundarstufe II (TIMSS/III) – Technische Grundlagen. In J. Baumert, W. Bos & R. Lehmann (Hrsg.), *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn: Bd. 1. Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit* (S. 31–84). Opladen: Leske + Budrich.
- Baumert, J., Köller, O., & Schnabel, K. U. (2000). Schulformen als differentielle Entwicklungsmilieus – eine ungehörige Fragestellung? In Gewerkschaft Erziehung und Wissenschaft (Hrsg.), *Messung sozialer Motivation. Eine Kontroverse* (S. 28–69). Frankfurt a.M.: Bildungs- und Förderungswerk der GEW (Schriftenreihe des Bildungs- und Förderungswerks der GEW).
- Baumert, J., Kunter, M., Brunner, M., Krauss, S., Blum, W., & Neubrand, M. (2004). Mathematikunterricht aus Sicht der PISA-Schülerinnen und -Schüler und ihrer Lehrkräfte. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Eds.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 314–354). Münster: Waxmann.
- Baumert, J., Schmitz, B., Sang, F., & Roeder, P. M. (1987). Zur Kompatibilität von Leistungsförderung und Divergenzminderung in Schulklassen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 19 (3), 249–265.
- Baumert, J., & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 323–407). Opladen: Leske + Budrich.
- Baumert, J., & Schümer, G. (2002). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb im nationalen Vergleich. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Die Länder der Bundesrepublik im Vergleich* (S. 159–202). Opladen: Leske + Budrich.
- Baumert, J., Stanat, P., & Demmrich, A. (2001). PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 15–68). Opladen: Leske + Budrich.
- Baumert, J., Trautwein, U., & Artelt, C. (2003). Schulumwelten – institutionelle Bedingungen des Lehrens und Lernens. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (S. 261–332). Opladen: Leske + Budrich.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107 (2), 238–246.
- Bentler, P. M. (1995). *EQS: A structural equations program manual*. Encino, CA: Multivariate Software, Inc.
- Berg, C. A. (2000). Intellectual development in adulthood. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 117–137). Cambridge, UK: Cambridge University Press.
- Bickley, P. G., Keith, T. Z., & Wolffe, L. M. (1995). The Three-Stratum Theory of cognitive abilities: Test of the structure of intelligence across the life span. *Intelligence*, 20, 309–328.
- Bloom, B. S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.
- Blum, W., Neubrand, M., Ehmke, T., Senkbeil, M., Jordan, A., Ulfig, F., & Carstensen, C. H. (2004). Mathematische Kompetenz. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 47–92). Münster: Waxmann.

- Bobertag, O., & Hylla, E. (1928). *Begabungsprüfung für den Übergang von der Grundschule zu den weiterführenden Schulen. Anleitung und Testheft*. Langensalza: Beltz.
- Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement*, 12 (3), 261–280.
- Boekaerts, M. (1996). Self-regulated learning at the junction of cognition and motivation. *European Psychologist*, 1 (2), 100–112.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110 (2), 305–314.
- Bond, L. (1989). The effects of special preparation on measures of scholastic ability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 429–444). New York: American Council on Education/Macmillan.
- Bourdieu, P. (1983). Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. In R. Kreckel (Hrsg.), *Soziale Ungleichheiten* (S. 183–198). Göttingen: Schwartz.
- Bortz, J., & Döring, N. (2002). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. New York: Springer.
- Bos, W., & Postlethwaite, T. N. (2001). Internationale Schulleistungsforschung. Ihre Entwicklungen und Folgen für die deutsche Bildungslandschaft. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 251–267). Weinheim: Beltz.
- Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, 53, 371–399.
- Brähler, E., Holling, H., Leutner, D., & Petermann, F. (Hrsg.). (2002). *Brückenkamp Handbuch psychologischer und pädagogischer Tests*. Göttingen: Hogrefe.
- Brody, N. (2000). History of theories and measurements of intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 16–33). Cambridge, UK: Cambridge University Press.
- Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie. Bd 3: Psychologie des Unterrichts und der Schule* (S. 177–212). Göttingen: Hogrefe.
- Bronfenbrenner, U., & Ceci, S. J. (1994). Nature-nurture reconceptualized in developmental perspective: A bioecological model. *Psychological Review*, 101, 568–586.
- Brophy, J. (2000). *Teaching*. Geneva: IBE.
- Brophy, J., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328–375). New York: Macmillan.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.
- Bruinsma, M. (2004). Motivation, cognitive processing and achievement in higher education. *Learning and Instruction*, 14, 549–568.
- Brunner, M., & Süß, H.-M. (2005). Analyzing the reliability of multidimensional measures: An example from intelligence research. *Educational & Psychological Measurement*, 65, 227–240.
- Carlstedt, B. (2001). Differentiation of cognitive abilities as a function of level of general intelligence: A latent variable approach. *Multivariate Behavioral Research*, 36, 589–609.
- Carpenter, T. P., & Lehrer, R. (1999). Teaching and learning mathematics with understanding. In E. Fennema & T. A. Romberg (Eds.), *Mathematics classrooms that promote understanding* (pp. 19–33). Mahwah, NJ: Erlbaum.
- Carroll, J. B. (1963). A model of school learning. *Teacher College Record*, 64 (8), 723–733.
- Carroll, J. B. (1966). Factors of verbal achievement. In A. Anastasi (Ed.), *Testing problems in perspective* (pp. 406–413). Washington, DC: American Council on Education.
- Carroll, J. B. (1974). The aptitude-achievement distinction: The case of foreign language aptitude and proficiency. In D. R. Green (Ed.), *The aptitude-achievement distinction* (pp. 286–303). Monterey, CA: CTB/McGraw-Hill.
- Carroll, J. B. (1989). The Carroll model: A 25-year retrospective and prospective view. *Educational Researcher*, 18 (1), 26–31.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factoranalytic studies*. New York: Cambridge University Press.
- Carroll, J. B. (1996). Mathematical abilities: Some results from factor analysis. In R. J. Sternberg & T. Ben-Zeev (Eds.), *The nature of mathematical thinking* (pp. 3–25). Mahwah, NJ: Erlbaum.
- Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports g and about ten broad factors. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5–22). New York: Pergamon.
- Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin*, 38, 592.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54 (1), 1–22.

- Cattell, R. B. (1987). *Intelligence: Its structure, growth, and action*. Amsterdam: Elsevier.
- Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology*, 27 (5), 703–722.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233–255.
- Clausen, M. (2002). *Unterrichtsqualität: eine Frage der Perspektive? Empirische Analysen zur Übereinstimmung, Konstrukt- und Kriteriumsvalidität*. Münster: Waxmann.
- Clary, T. A., Humphreys, L. G., Kendrick, S. A., & Wesman, A. (1975). Educational uses of tests with disadvantaged students. *American Psychologist*, 30, 15–41.
- Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research*, 18 (1), 115–126.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112 (1), 155–159.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94, 95–120.
- Colom, R., Rebollo, I., Palacios, A., Juan-Espinosa, M., & Kyllonen, P. C. (2004). Working memory is (almost) perfectly predicted by g. *Intelligence*, 32, 277–296.
- Cooley, W. W., & Lohnes, P. R. (1976). *Evaluation research in education*. New York: Irvington.
- Corno, L., Cronbach, L. J., Kupermintz, H., Lohman, D. F., Mandinach, E. B., Porteus, A. W., & Talbert, J. E. (2002). *Remaking the concept of aptitude: Extending the legacy of Richard E. Snow*. Mahwah, NJ: Erlbaum.
- Corno, L., & Snow, R. E. (1986). Adapting teaching to individual differences among learners. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 605–629). New York: Macmillan.
- Covington, M. V. (2000). Goal theory, motivation, and school achievement: An integrative review. *Annual Review of Psychology*, 51, 171–200.
- Creemers, B. P. M. (1994). *The effective classroom*. London: Cassell.
- Cronbach, L. J. (1984). *Essentials of psychological testing* (4th ed.). New York: Harper & Row.
- Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York: Irvington.
- Daniels, Z. (2005). *Differenzierung von Interessen im Jugendalter*. Unveröff. Dissertation, Universität Potsdam.
- Dark, V. J., & Benbow, C. P. (1990). Enhanced problem translation and short-term memory: Components of mathematical talent. *Journal of Educational Psychology*, 82 (3), 420–429.
- Dark, V. J., & Benbow, C. P. (1991). Differential enhancement of working memory with mathematical versus verbal precocity. *Journal of Educational Psychology*, 83, 48–60.
- De Corte, E., Greer, B., & Verschaffel, L. (1996). Mathematics teaching and learning. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 491–549). New York: Macmillan.
- Deary, I. J., Egan, V., Gibson, G. J., Austin, E. J., Brand, C. R., & Kellaghan, T. (1996). Intelligence and the differentiation hypothesis. *Intelligence*, 23, 105–132.
- Deary, I. J., & Pagliari, C. (1991). The strength of g at different levels of ability: Have Detterman and Daniel rediscovered Spearman's "law of diminishing returns"? *Intelligence*, 15, 247–250.
- Department of Defense. (1984). *Test manual for the Armed Services Vocational Aptitude Battery* (DOD 1304.12AA). North Chicago: U.S. Military Entrance Processing Command.
- DeShon, R. P. (1998). A cautionary note on measurement error in corrections in structural equation models. *Psychological Methods*, 3 (4), 412–423.
- Detterman, D. K., & Daniel, M. H. (1989). Correlations of mental tests with each other and with cognitive variables are highest for low IQ groups. *Intelligence*, 13, 349–359.
- Deutsches PISA-Konsortium. (2000). *Schülerleistungen im internationalen Vergleich. Eine neue Rahmenkonzeption für die Erfassung von Wissen und Fähigkeiten*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Dixon, R. C., Carnine, D. W., Lee, D.-S., Wallin, J., & Chard, D. (1998). *Report to the California State Board of Education and addendum to principal Report: Review of high quality experimental mathematics research*. <<http://www.idea.uoregon.edu/~ncite/documents/math/math.html>> (13/5/2003)
- Döbert, H., & Geißler, G. (2000). *Schulleistung in der DDR. Das System der Leistungsentwicklung, Leistungssicherung und Leistungsmessung*. Frankfurt a.M.: Lang.
- Dochy, F. J. R. C., & Alexander, P. A. (1995). Mapping prior knowledge: A framework for discussion among researchers. *European Journal of Psychology of Education*, 10, 225–242.
- Du Bois, P. H. (Ed.). (1969). *Invitational conference on testing problems*. Princeton, NJ: Educational Testing Service.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53, 109–132.
- Educational Research Centre. (1986). *Differential aptitude tests: Form T, Manual*. Dublin: Educational Research Centre.

- Einsiedler, W. (1997). Unterrichtsqualität und Leistungsentwicklung: Literaturüberblick. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 225–240). Weinheim: Beltz.
- Ekstrom, R. B., French, J. W., & Harman, H. H. (1979). Cognitive factors: Their identification and replication. *Multivariate Behavioral Research Monographs*, 79 (2).
- Enders, C. K. (2001). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling*, 8 (1), 128–141.
- Enders, C. K. (2003). Using the expectation maximization algorithm to estimate coefficient alpha for scales with item-level missing data. *Psychological Methods*, 8, 322–337.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, 128 (3), 309–331.
- Ferguson, G. A. (1954). On learning and human ability. *Canadian Journal of Psychology*, 8, 95–112.
- Ferguson, G. A. (1956). On transfer and the abilities of man. *Canadian Journal of Psychology*, 10, 121–130.
- Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler intelligence scales and Gf-Gc theory: A contemporary approach to interpretation*. Boston: Allyn & Bacon.
- Floden, R. E. (2002). The measurement of opportunity to learn. In National Research Council (Ed.), *Methodological advances in cross-sectional surveys of educational achievement* (pp. 231–266). Washington, DC: National Academy Press.
- Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, 9, 466–491.
- Fogarty, G. J., & Stankov, L. (1995). Challenging the “law of diminishing returns”. *Intelligence*, 21, 157–174.
- Freudenthal, H. (1977). *Mathematik als pädagogische Aufgabe* (2 Bde.). Stuttgart: Klett.
- Ganzeboom, H. B. G., de Graaf, P. M., Treiman, D. J., & de Leeuw, J. (1992). A standard international socioeconomic index of occupational status. *Social Science Research*, 21, 1–56.
- Garrett, H. E. (1946). A developmental theory of intelligence. *American Psychologist*, 1, 372–378.
- Geary, D. C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin*, 114, 345–362.
- Geary, D. C. (1994). *Children’s mathematical development: Research and practical applications*. Washington, DC: American Psychological Association.
- Geary, D. C. (1995). Reflections of evolution and culture in children’s cognition. Implications for mathematical development and instruction. *American Psychologist*, 50, 24–37.
- Geary, D. C. (1996). Sexual selection and sex differences in mathematical abilities. *Behavioral and Brain Sciences*, 19, 229–284.
- Geary, D. C. (2000). From infancy to adulthood: The development of numerical abilities. *European Child & Adolescent Psychiatry*, 9, 11–16.
- Geary, D. C. (2005). *The origin of mind: Evolution of brain, cognition, and general intelligence*. Washington, DC: American Psychological Association.
- Geary, D. C., & Widaman, K. F. (1992). Numerical cognition: On the convergence of componential and psychometric models. *Intelligence*, 16, 47–80.
- Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and the animate-inanimate distinction as examples. *Cognitive Science*, 14, 79–106.
- Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist*, 39, 93–104.
- Glöckner-Rist, A., & Hoijtink, H. (2003). The best of both worlds: Factor analysis of dichotomous data using item response theory and structural equation modeling. *Structural Equation Modeling*, 10, 544–565.
- Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history and bibliography. *Intelligence*, 24, 13–23.
- Graham, J. W., Cumsille, P. E., & Elek-Fisk, E. (2003). Methods for handling missing data. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology: Vol. 2. Research methods in psychology* (pp. 87–114). New York: Wiley.
- Green, D. R. (Ed.). (1974). *The aptitude-achievement distinction*. Monterey, CA: CTB/McGraw-Hill.
- Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 15–46). New York: Macmillan.
- Greeno, J. G., Smith, D. R., & Moore, J. L. (1993). Transfer of situated learning. In D. K. Detterman & R. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction* (pp. 99–167). Norwood, NJ: Ablex.
- Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, 6 (4), 430–450.
- Gruehn, S. (2000). *Unterricht und schulisches Lernen: Schüler als Quellen der Unterrichtsbeschreibung*. Münster: Waxmann.
- Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.
- Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, 8, 179–203.



- Gustafsson, J. E. (1994). Hierarchical models of intelligence and educational achievement. In A. Demetriou & A. Efklides (Eds.), *Intelligence, mind, and reasoning: structure and development* (pp. 45–73). Amsterdam: Elsevier.
- Gustafsson, J., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, 28, 407–434.
- Gustafsson, J. E., & Snow, R. E. (1997). Ability profiles. In R. F. Dillon (Ed.), *Handbook on testing* (pp. 107–135). Westport, CT: Greenwood Press/Greenwood Publishing Group, Inc.
- Gustafsson, J., & Undheim, J. O. (1996). Individual differences in cognitive functions. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 186–242). New York: Macmillan.
- Guttman, L. (1965). A faceted definition of intelligence. In R. R. Eifermann (Ed.), *Studies in psychology* (Vol. XIV, pp. 166–229). Jerusalem: The Hebrew University.
- Guttman, L., & Levy, S. (1991). Two structural laws for intelligence tests. *Intelligence*, 15, 79–103.
- Haertel, G. D., Walberg, H. J., & Weinstein, T. (1983). Psychological models of educational performance: A theoretical synthesis of constructs. *Review of Educational Research*, 53 (1), 75–91.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Hansford, B. C., & Hattie, J. A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research*, 52, 123–142.
- Harnishfeger, A., & Wiley, D. E. (1977). Kernkonzepte des Schullernens. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 9 (3), 207–228.
- Hartmann, P., & Reuter, M. (2006). Spearman's "law of diminishing returns" tested with two methods. *Intelligence*, 34, 47–62.
- Heckhausen, H. (1989). *Motivation und Handeln* (2., völlig überarb. u. erg. Aufl.). Berlin: Springer.
- Hegarty, M., Mayer, R. E., & Green, C. E. (1992). Comprehension of arithmetic word problems: Evidence from student's eye fixations. *Journal of Educational Psychology*, 84 (1), 76–84.
- Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology*, 87 (1), 18–32.
- Heller, K. A., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision. Manual*. Göttingen: Hogrefe.
- Helmke, A. (2003). *Unterrichtsqualität erfassen, bewerten, verbessern*. Seelze: Kallmeyersche Verlagsbuchhandlung.
- Helmke, A., Schneider, W., & Weinert, F. E. (1986). Quality of instruction and classroom learning outcomes: The German contribution to the IEA classroom environment study. *Teaching & Teacher Education*, 2 (1), 1–18.
- Helmke, A., & Schrader, F.-W. (1987). Interactional effects of instructional quality and teacher judgement accuracy on achievement. *Teaching & Teacher Education*, 3 (2), 91–98.
- 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, 624–637.
- Helmke, A., & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie: Bd. 3. Psychologie des Unterrichts und der Schule* (S. 71–176). Göttingen: Hogrefe.
- Hiebert, J., & Carpenter, T. (1992). Learning and teaching with understanding. In D. A. Grouws (Ed.), *NCTM Handbook of research on mathematics teaching and learning* (pp. 65–97). New York: Macmillan.
- Horn, J. (1988). Thinking about human abilities. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 645–685). New York: Plenum.
- Horn, J. L. (1994). Theory of fluid and crystallized intelligence. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 443–451). New York: Macmillan.
- Horn, J. L., & Knapp, J. R. (1973). On the subjective character of the empirical base of Guilford's Structure-of-Intellect model. *Psychological Bulletin*, 80, 33–43.
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18, 117–144.
- Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53–91). New York: The Guilford Press.
- Hosenfeld, I., Köller, O., & Baumert, J. (1999). Why sex differences in mathematics achievement disappear in German secondary schools: A reanalysis of the German TIMSS-data. *Studies in Educational Evaluation*, 25, 143–161.
- Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3 (4), 424–453.
- Humphreys, L. G. (1984). General intelligence. In C. R. Reynolds & R. T. Brown (Eds.), *Perspectives on bias in mental testing* (pp. 221–248). New York: Plenum Press.
- Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, 107, 139–155.

- Ingenkamp, K. (1962). *Die deutschen Schulleistungstests*. Weinheim: Beltz.
- Ingenkamp, K. (1968). Die Anwendung von Schultests in Deutschland. In K. Ingenkamp & T. Marsolek (Hrsg.), *Möglichkeiten und Grenzen der Testanwendung in der Schule* (S. 83–94). Weinheim: Beltz.
- International Labor Office. (1990). *International standard classification of occupations: ISCO-88*. Geneva: International Labor Office.
- Jäger, A. O. (1967). *Dimensionen der Intelligenz*. Göttingen: Hogrefe.
- Jäger, A. O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen. Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells. *Diagnostica*, 28, 195–226.
- Jäger, A. O. (1984). Intelligenzstrukturforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven. *Psychologische Rundschau*, 35, 21–35.
- Jäger, A. O., Süß, H.-M., & Beauducel, A. (1997). *Berliner Intelligenzstruktur – Test, Form 4*. Göttingen: Hogrefe.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport: Praeger.
- Jensen, A. R. (2003). Regularities in Spearman's law of diminishing returns. *Intelligence*, 31, 95–105.
- Johnson, S. T. (1994). Scholastic Assessment Tests (SAT). In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 956–960). New York: Macmillan.
- Jordan, A., Ross, N., Krauss, S., Baumert, J., Blum, W., Neubrand, M., Löwen, K., Brunner, M., & Kunter, M. (2006). *Ein Klassifikationsschema für Mathematikaufgaben*. Berlin: Max-Planck-Institut für Bildungsforschung (Materialien aus der Bildungsforschung 81).
- Jöreskog, K. G., Sörbom, D., Du Toit, S., & Du Toit, M. (1999). *LISREL 8: New statistical features*. Chicago: Scientific Software International.
- Juan-Espinosa, M., Garcia, L. F., Escorial, S., Rebollo, I., Colom, R., & Abad, F. J. (2002). Age dedifferentiation hypothesis: Evidence from the WAIS III. *Intelligence*, 30, 395–408.
- Kahl, R., & Spiewak, M. (2005). *Nur bedingt wissenschaftlich*. <<http://www.hermes.zeit.de/pdf/archiv/2005/11/B-Erziehungswissenschaften.pdf>> (05/13/2005)
- Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, 131, 66–71.
- Kaplan, D. (1990). Evaluating and modifying covariance structure models: A review and recommendation. *Multivariate Behavioral Research*, 25 (2), 137–155.
- Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, 49 (4), 294–303.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge University Press.
- Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review*, 92 (1), 109–129.
- Klieme, E. (2000). Fachleistungen im voruniversitären Mathematik- und Physikunterricht: Theoretische Grundlagen, Kompetenzstufen und Unterrichtsschwerpunkte. In J. Baumert, W. Bos & R. Lehmann (Hrsg.), *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn: Bd. 2. Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe* (S. 57–128). Opladen: Leske + Budrich.
- Klieme, E., & Baumert, J. (2001). Identifying national cultures of mathematics education: Analysis of cognitive demands and differential item functioning in TIMSS. *European Journal of Psychology in Education*, 16 (3), 385–402.
- Klieme, E., Neubrand, M., & Lüdtke, O. (2001). Mathematische Grundbildung: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 139–190). Opladen: Leske + Budrich.
- Klieme, E., & Rakoczy, K. (2003). Unterrichtsqualität aus Schülerperspektive: Kulturspezifische Profile, regionale Unterschiede und Zusammenhänge mit Effekten von Unterricht. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (S. 333–360). Opladen: Leske + Budrich.
- Klieme, E., Schümer, G., & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: “Aufgabenkultur” und Unterrichtsgestaltung. In E. Klieme & J. Baumert (Hrsg.), *TIMSS – Impulse für Schule und Unterricht. Forschungsbefunde, Reforminitiativen, Praxisberichte und Video-Dokumente* (S. 43–57). Bonn: Bundesministerium für Bildung und Forschung.
- Köller, O. (1998). *Zielorientierungen und schulisches Lernen*. Münster: Waxmann.
- Köller, O. (2003). Gesamtschule – Erweiterung statt Alternative. In K. S. Cortina, J. Baumert, A. Leschinsky, K. U. Mayer & L. Trommer (Hrsg.), *Das Bildungswesen in der Bundesrepublik Deutschland. Strukturen und Entwicklungen im Überblick* (S. 458–486). Reinbek: Rowohlt.

- Köller, O., & Baumert, J. (2001). Leistungsgruppierungen in der Sekundarstufe I. Ihre Konsequenzen für die Mathematikleistung und das mathematische Selbstkonzept der Begabung. *Zeitschrift für pädagogische Psychologie*, 15, 99–110.
- Köller, O., & Baumert, J. (2002). Entwicklung schulischer Leistungen. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (5., vollst. überarb. Aufl., S. 756–786). Weinheim: Beltz.
- Köller, O., Baumert, J., & Schnabel, K. U. (2001). Does interest matter? The relationship between academic interest and achievement in mathematics. *Journal for Research in Mathematics Education*, 32 (5), 448–470.
- Köller, O., Baumert, J., & Schnabel, K. U. (2003). Secondary school as a constraint for adolescent development. In U. M. Staudinger & U. Lindenberger (Eds.), *Understanding human development: Dialogues with lifespan psychology* (pp. 449–461). Boston: Kluwer.
- Köller, O., Daniels, Z., & Baumert, J. (2000, April 24–28). *Multiple frames of reference, academic interests, and coursework selection in upper secondary schools*. Paper presented at the AERA Annual Meeting, New Orleans.
- Köller, O., Schnabel, K., & Baumert, J. (2000). Der Einfluß der Leistungsstärke von Schulen auf das fachspezifische Selbstkonzept der Begabung und das Interesse. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 32, 70–80.
- Köller, O., Watermann, R., & Baumert, J. (2001). Skalierung der Leistungstests in PISA. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 517–524). Opladen: Leske + Budrich.
- Kovas, Y., Harlaar, N., Petrill, S. A., & Plomin, R. (2005). ‘Generalist genes’ and mathematics in 7-year-old twins. *Intelligence*, 33, 473–489.
- Krapp, A. (1998). Entwicklung und Förderung von Interessen im Unterricht. *Psychologie in Erziehung und Unterricht*, 44, 185–201.
- Krapp, A. (2002). Structural and dynamic aspects of interest development: Theoretical considerations from an ontogenetic perspective. *Learning and Instruction*, 12, 383–409.
- Krauss, S., Kunter, M., Brunner, M., Baumert, J., Blum, W., Neubrand, M., Jordan, A., & Löwen, K. (2004). COACTIV: Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz. In J. Doll & M. Prenzel (Hrsg.), *Die Bildungsqualität von Schule: Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung* (S. 31–53). Münster: Waxmann.
- Kulik, J. A., Kulik, C.-L., & Bangert-Drowns, R. L. (1984). Effects of practice on aptitude and achievement test scores. *American Educational Research Journal*, 21 (2), 435–447.
- Kunter, M. (2005). *Multiple Ziele im Mathematikunterricht*. Münster: Waxmann.
- Kunter, M., Schümer, G., Artelt, C., Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J., & Weiß, M. (2002). *PISA 2000: Dokumentation der Erhebungsinstrumente*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Kupermintz, H., Ennis, M. M., Hamilton, L. S., Talbert, J. E., & Snow, R. E. (1995). Enhancing the validity and usefulness of large-scale educational assessments: I. NELS: 88 mathematics achievement. *American Educational Research Journal*, 32 (3), 525–554.
- Kupermintz, H., & Snow, R. E. (1997). Enhancing the validity and usefulness of large-scale educational assessments: III. NELS: 88 mathematics achievement to 12th grade. *American Educational Research Journal*, 34 (1), 124–150.
- Kyllonen, P. C. (1994). Aptitude testing inspired by information processing: A test of the four-sources model. *The Journal of General Psychology*, 120, 375–405.
- Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, 14, 389–433.
- Legree, P. J., Pifer, M. E., & Grafton, F. C. (1996). Correlations among cognitive abilities are lower for higher ability groups. *Intelligence*, 23, 45–57.
- Lehmann, R. H., Hunger, S., Ivanov, S., & Gänsfuß, R. (2004). *LAU 11. Aspekte der Lernausgangslage und der Lernentwicklung. Klassenstufe 11*. <<http://www.hamburger-bildungsserver.de/schulentwicklung/lau/lau11/kap.1-2.pdf>> (15/06/2005)
- Leutner, D. (2001). Pädagogisch-psychologische Diagnostik. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (2. Aufl., S. 521–530). Weinheim: Psychologie Verlags Union.
- Leutner, D., Klieme, E., Meyer, K., & Wirth, J. (2004). Problemlösen. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 147–175). Münster: Waxmann.

- Li, S.-C., Lindenberger, U., Hommel, B., Aschersleben, G., Prinz, W., & Baltes, P. B. (2004). Transformations in the couplings among intellectual abilities and constituent cognitive processes across the life span. *Psychological Science, 15*, 155–163.
- Li, S.-C., & Schmiedek, F. (2001). Intelligence: Central conceptions and psychometric models. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (pp. 7635–7641). Amsterdam: Elsevier.
- Lind, D., & Knoche, N. (2004). Eine differenzielle Itemanalyse zu den Faktoren Bildungsgang und Geschlecht. In M. Neubrand (Hrsg.), *Mathematische Kompetenzen von Schülerinnen und Schülern in Deutschland. Vertiefende Analysen im Rahmen von PISA 2000* (S. 73–86). Wiesbaden: VS Verlag für Sozialwissenschaften.
- van der Linden, W. J. (2000). Optimal assembly of tests with item sets. *Applied Psychological Measurement, 24* (3), 225–240.
- van der Linden, W. J., & Hambleton, R. K. (Eds.). (1997). *Handbook of modern item response theory*. New York: Springer.
- Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research, 32*, 53–76.
- Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When “good” indicators are bad and “bad” indicators are good. *Psychological Methods, 4* (2), 192–211.
- Lohman, D. F. (1994). Component scores as residual variation (or why the intercept correlates best). *Intelligence, 19*, 1–11.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Lubinski, D. (2000). Scientific and social significance of assessing individual differences: “Sinking shafts at a few critical points”. *Annual Review of Psychology, 51*, 405–444.
- Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman’s (1904) “General intelligence’, objectively determined and measured”. *Journal of Personality and Social Psychology, 86*, 96–111.
- Lubinski, D., Webb, R. M., Morelock, M. J., & Benbow, C. P. (2001). Top 1 in 10,000: A 10-year follow-up of the profoundly gifted. *Journal of Applied Psychology, 86*, 718–729.
- Lubke, G. H., Dolan, C. V., Kelderman, H., & Mellenbergh, G. J. (2003). On the relationship between sources of within- and between-group differences and measurement invariance in the common factor model. *Intelligence, 31*, 543–566.
- 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.
- Ma, X., & Kishor, N. (1997). Attitude toward self, social factors, and achievement in mathematics: A meta-analytic review. *Educational Psychology Review, 9*, 89–120.
- Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal, 23*, 129–149.
- Marsh, H. W. (1989). Confirmatory factor analysis of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement, 13* (4), 335–361.
- Marsh, H. W. (1990). *Self Description Questionnaire (SDQ) II: A theoretical and empirical basis for the measurement of multiple dimensions of adolescent self-concept: An interim test manual and a research monograph*. San Antonio, TX: The Psychological Corporation.
- Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting of cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler’s (1999) findings. *Structural Equation Modeling, 11*, 320–341.
- Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development, 76*, 397–416.
- Marsh, H. W., & Yeung, A. S. (1997). Causal effects of academic self-concept on academic achievement: Structural equation models of longitudinal data. *Journal of Educational Psychology, 89*, 41–45.
- Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence, 7*, 107–127.
- Mayer, R. E. (1981). Frequency norms and structural analysis of algebra story problems into families, categories, and templates. *Instructional science, 10*, 135–175.
- Mayer, R. E. (1985). Mathematical ability. In R. J. Sternberg (Ed.), *Human abilities: An information-processing approach* (pp. 127–150). New York: Freeman.
- Mayer, R. E. (1992). *Thinking, problem solving, cognition*. New York: Freeman.
- Mayer, R. E. (1994). Mathematical ability. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 688–691). New York: Macmillan.
- Mayer, R. E. (2000). Intelligence and education. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 519–533). Cambridge, UK: Cambridge University Press.
- Mayer, R. E. (2004). Teaching of subject matter. *Annual Review of Psychology, 55*, 715–744.

- Mayer, R. E., & Hegarty, M. (1996). The process of understanding mathematical problems. In R. J. Sternberg & T. Ben-Zeev (Eds.), *The nature of mathematical thinking* (pp. 29–53). Mahwah, NJ: Erlbaum.
- Mayer, R. E., & Wittrock, M. C. (1996). Problem-solving and transfer. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 47–62). New York: Macmillan.
- McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, *34*, 100–117.
- McDonald, R. P. (1997). Normal-ogive multidimensional model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 258–271). New York: Springer.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.
- McDonald, R. P. (2000). A basis for multidimensional item response theory. *Applied Psychological Measurement*, *24* (2), 99–114.
- McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151–179). New York: The Guilford Press.
- McGrew, K. S., & Hessler, G. L. (1995). The relationship between the WJ-R Gf-Gc cognitive clusters and mathematics achievement across the life-span. *Journal of Psychoeducational Assessment*, *13*, 21–38.
- Messick, S. (1984). The psychology of educational measurement. *Journal of Educational Measurement*, *21*, 215–237.
- Millman, J., & Greene, J. (1989). The specification and development of tests of achievement and ability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 335–366). New York: American Council on Education and Macmillan Publishing Company.
- Millsap, R. E., & Kwok, O. M. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods*, *9*, 93–115.
- Mislevy, R. J., Beaton, A. E., Kaplan, B., & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, *29* (2), 133–161.
- Mitchell, M. (1993). Situational interest: Its multifaceted structure in the secondary school mathematics classroom. *Journal of Educational Psychology*, *85* (3), 424–436.
- Möller, J., & Köller, O. (2004). Die Genese akademischer Selbstkonzepte: Effekte dimensionaler und sozialer Vergleiche. *Psychologische Rundschau*, *55*, 19–27.
- Mulaik, S. A., & Quartetti, D. A. (1997). First order or higher order general factor? *Structural Equation Modeling*, *4* (3), 193–211.
- Mullis, I. V. S., Martin, M. O., Beaton, A. E., Gonzales, E. J., Kelly, D. L., & Smith, T. A. (1998). *Mathematics and science achievement in the final year of secondary school: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: IEA-TIMSS.
- Muraki, E., & Bock, R. D. (1996). Parscale (Version 4). Lincolnwood, IL: Scientific Software International.
- Muthén, B. O., Kao, C., & Burstein, L. (1991). Instructionally sensitive psychometrics: Application of new IRT-based detection technique to mathematics achievement test items. *Journal of Educational Measurement*, *28*, 1–22.
- Muthén, B. O., Khoo, S.-T., & Goff, G. N. (1997). *Multidimensional description of subgroup differences in mathematics achievement data from the 1992 National Assessment of Educational Progress*. Los Angeles: CRESST/University of California.
- Muthén, L. K., & Muthén, B. O. (1998–2004a). *Mplus technical appendices*. Los Angeles: Muthén & Muthén.
- Muthén, L. K., & Muthén, B. O. (1998–2004b). *Mplus user's guide: Third edition*. Los Angeles: Muthén & Muthén.
- Nagy, G., Trautwein, U., Baumert, J., Köller, O., & Garrett, J. (in press). Gender and course selection in upper secondary education: Effects of academic self-concept and intrinsic value. *Educational Research and Evaluation*.
- Nathan, M. J., Kintsch, W., & Young, E. (1992). A theory of algebra-word-problem comprehension and its implications for the design of learning environments. *Cognition and Instruction*, *9*, 329–389.
- NCES – National Center for Education Statistics. (2005). *Frequently asked questions*. <<http://www.nces.ed.gov/nationsreportcard/faq.asp#ques1>> (01/11/2005)
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics (NCTM).
- Neisser, U., Boodoo, G., Bouchard, T. J. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*, 77–101.
- Nesselroade, J. R., & Thompson, W. W. (1995). Selection and related threats to group comparisons: An example comparing factorial structures of higher and lower ability groups of adult twins. *Psychological Bulletin*, *117*, 271–284.
- Neubrand, M., Biehler, R., Blum, W., Cohors-Fresenborg, E., Flade, L., Knoche, N., Lind, D., Löding, W., Möller, G., & Wynands, A. (Deutsche PISA-2000-Expertengruppe Mathematik). (2001). Grundlagen der Ergänzung des internationalen PISA-Mathematik-Tests in der deutschen Zusatzerhebung. *Zentralblatt für Didaktik der Mathematik*, *33* (2), 45–59.

- Neubrand, M., Klieme, E., Lüdtke, O., & Neubrand, J. (2002). Kompetenzstufen und Schwierigkeitsmodelle für den PISA-Test zur mathematischen Grundbildung. *Unterrichtswissenschaft, 30*, 100–119.
- Oberauer, K., Schulze, R., Wilhelm, O., & Süß, H.-M. (2005). Working memory and intelligence – their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*, 61–65.
- Pekrun, R., & Schiefele, U. (1997). Emotions- und motivationspsychologische Bedingungen der Lernleistung. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie: Bd. 3. Psychologie des Lernens und der Instruktion* (S. 153–180). Göttingen: Hogrefe.
- Pellegrino, J. W., Baxter, G. P., & Glaser, R. (1999). Addressing the “Two disciplines” problem: Linking theories of cognition and learning with assessment and instructional practice. *Review of Research in Education, 24*, 307–354.
- Pintrich, P. R. (2003). Motivation and classroom learning. In W. M. Reynolds & G. E. Miller (Eds.), *Handbook of psychology: Vol. 7. Educational psychology* (pp. 103–122). Hoboken, NJ: Wiley.
- Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P. (2001). *Behavioral genetics* (4th ed.). New York: Worth Publishers.
- Plomin, R., & Petrill, S. A. (1997). Genetics and intelligence: What’s new? *Intelligence, 24*, 53–77.
- Prenzel, M., Drechsel, B., Carstensen, C. H., & Ramm, G. (2004). PISA 2003 – Eine Einführung. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 13–46). Münster: Waxmann.
- Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Psychological Methods, 8*(7), 517–529.
- Raudenbush, S. W., Rowan, B., & Cheong, Y. F. (1993). Higher order instructional goals in secondary schools: Class, teacher, and school influences. *American Educational Research Journal, 30* (3), 523–553.
- Reeve, C. L. (2004). Differential ability antecedents of general and specific dimensions of declarative knowledge: More than g. *Intelligence, 32*, 621–652.
- Reinert, G. (1970). Comparative factor analytic studies of intelligence throughout the human life-span. In L. R. Goulet & P. B. Baltes (Eds.), *Life-span development and behavior: Research and theory* (pp. 467–484). New York: Academic Press.
- Reinert, G., Baltes, P. B., & Schmidt, L. R. (1965). Faktoranalytische Untersuchungen zur Differenzierungshypothese der Intelligenz: Die Leistungsdifferenzierungshypothese. *Psychologische Forschung, 28*, 246–300.
- Renkl, A. (1996). Vorwissen und Schulleistung. In J. Möller & O. Köller (Hrsg.), *Emotionen, Kognitionen und Schulleistung* (S. 175–190). Weinheim: Beltz.
- Renkl, A., & Stern, E. (1994). Die Bedeutung von kognitiven Eingangsvoraussetzungen und schulischen Lerngelegenheiten für das Lösen von einfachen und komplexen Textaufgaben. *Zeitschrift für pädagogische Psychologie, 8* (1), 27–39.
- Resnick, L. B. (1989). Developing mathematical knowledge. *American Psychologist, 44* (2), 162–169.
- Rheinberg, F. (1996). Von der Lernmotivation zur Lernleistung: Was liegt dazwischen? In J. Möller & O. Köller (Hrsg.), *Emotionen, Kognitionen und Schulleistung* (S. 23–52). Weinheim: Beltz.
- Rindskopf, D. (2000). Plausible rival hypotheses in measurement, design and scientific theory. In L. Bickman (Ed.), *Research design: Donald Campbell’s legacy* (pp. 1–12). Thousand Oaks, CA: Sage.
- Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology, 91*, 175–189.
- Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93* (2), 346–362.
- Robitaille, D. F., & Garden, A. (Eds.). (1989). *The IEA study of mathematics II: Contexts and outcomes of school mathematics*. Oxford: Pergamon Press.
- Rost, D. H. (2001). Hochbegabung. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (2. Aufl., S. 239–248). Weinheim: Psychologie Verlags Union.
- Rost, J. (2004a). *Lehrbuch Testtheorie – Testkonstruktion* (2., vollst. überarb. und erw. Aufl.). Bern: Huber.
- Rost, J. (2004b). *Zur Psychometrie von PISA 2003*. Kiel: Institut für Pädagogik der Naturwissenschaften.
- Rost, D. H., Dickhäuser, O., Sparfeldt, J. R., & Schilling, S. R. (2004). Fachspezifische Selbstkonzepte und Schulleistungen im dimensional Vergleich. Eine versuchsplanerische Überprüfung des I/E-Modells. *Zeitschrift für pädagogische Psychologie, 18*, 43–52.
- Royce, J. R. (1973). The conceptual framework for a multi-factor theory of individuality. In J. R. Royce (Ed.), *Multivariate analysis and psychological theory* (pp. 305–408). London: Academic Press.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- Ryan, R. M. (1995). Psychological needs and the facilitation of integrative processes. *Journal of Personality, 63*, 397–427.

- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66, 507–514.
- Schafer, J. L. (2000). *NORM for Windows 95/98/NT (Version 2.03)*. <<http://www.stat.psu.edu/~jls/misoftwa.html>> (11/2/2004)
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7 (2), 147–177.
- Schaie, K. W. (1994). The course of adult intellectual development. *American Psychologist*, 49, 304–313.
- Schaie, K. W., Maitland, S. B., Willis, S. L., & Intrieri, R. C. (1998). Longitudinal invariance of adult psychometric ability factor structures across 7 years. *Psychology and Aging*, 13, 8–20.
- Schaie, K. W., Willis, S. L., Jay, G., & Chipuer, H. (1989). Structural invariance of cognitive abilities across the adult life span: A cross-sectional study. *Developmental Psychology*, 25, 652–662.
- Schatz Koehler, M., & Grouws, D. A. (1992). Mathematics teaching practices and their effects. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 115–126). New York: Macmillan.
- 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.
- Schiefele, U., & Pekrun, R. (1997). Psychologische Modelle des fremdgesteuerten und selbstgesteuerten Lernens. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie: Bd. 3. Psychologie des Lernens und der Instruktion* (S. 249–278). Göttingen: Hogrefe.
- Schilling, S. R., Sparfeldt, J. R., & Rost, D. H. (2004). Wie generell ist das Modell? Analysen zum Geltungsbereich des “Internal/External Frame of Reference”-Modells. *Zeitschrift für pädagogische Psychologie*, 18, 221–230.
- Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22, 53–61.
- Schmidt, W. H., McKnight, C. C., Valverde, G. A., Houang, R. T., & Wiley, D. E. (1996). *Many visions, many aims: A cross-national investigation of curricular intentions in school mathematics*. Dordrecht: Kluwer.
- Schmiedek, F. (2003). *The structure of cognitive abilities in old and very old age: On the importance of specific group factors in a differentiated factor space*. Berlin: Freie Universität Berlin.
- Schmiedek, F., & Li, S.-C. (2004). Toward an alternative representation for disentangling age-associated differences in general and specific cognitive abilities. *Psychology and Aging*, 19, 40–56.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8, 350–353.
- Schneider, W., & Stefanek, J. (2004). Entwicklungsveränderungen allgemeiner kognitiver Fähigkeiten und schulbezogener Fertigkeiten im Kindes- und Jugendalter. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 36, 147–159.
- Schoenfeld, A. H. (1985). *Mathematical problem solving*. Orlando, FL: Academic Press.
- Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. A. Grouws (Ed.), *NCTM Handbook of research on mathematics teaching and learning* (pp. 334–370). New York: Macmillan.
- Schrader, F.-W. (1997). Lern- und Leistungsdiagnostik im Unterricht. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie: Bd. 3. Psychologie des Unterrichts und der Schule* (S. 659–700). Göttingen: Hogrefe.
- Schrader, F.-W., & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *Empirische Pädagogik*, 1 (1), 27–52.
- Schrader, F.-W., & Helmke, A. (2001). Alltägliche Leistungsbeurteilungen durch Lehrer. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 45–58). Weinheim: Beltz.
- Schulze, R. (2005). Modeling structures of intelligence. In O. Wilhelm & R. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 241–264). Mahwah, NJ: Erlbaum.
- Schupp, H. (1988). Anwendungsorientierter Mathematikunterricht in der Sekundarstufe I zwischen Tradition und neuen Impulsen. *Der Mathematikunterricht*, 34, 5–16.
- Slavin, R. E. (1994). Quality, appropriateness, incentive, and time: A model of instructional effectiveness. *International Journal of Educational Research*, 21, 141–157.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.
- Snow, R. E. (1980). Aptitude and achievement. *New Directions for Testing and Measurement*, 5, 39–59.
- Snow, R. E. (1989). Aptitude-treatment interaction as a framework for research on individual differences in learning. In P. L. Ackerman, R. J. Sternberg & R. Glaser (Eds.), *Learning and individual differences: Advances in theory and research* (pp. 13–60). New York: Freeman.
- Snow, R. E., Corno, L., & Jackson, D., III. (1996). Individual differences in affective and cognitive functions. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 243–310). New York: Macmillan.
- Snow, R. E., Kyllonen, P. C., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 47–103). Hillsdale, NJ: Erlbaum.

- Snow, R. E., & Lohman, D. F. (1984). Toward a theory of cognitive aptitude for learning from instruction. *Journal of Educational Psychology*, 76, 347–376.
- Snow, R. E., & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 263–331). New York: American Council on Education and Macmillan Publishing Company.
- Snow, R. E., & Swanson, J. (1992). Instructional psychology: Aptitude, adaption, and assessment. *Annual Review of Psychology*, 43, 583–626.
- Snyderman, M., & Rothman, S. (1987). Survey of expert opinion on intelligence and aptitude testing. *American Psychologist*, 42, 137–144.
- Spada, H., & Wichmann, S. (1996). Kognitive Determinanten der Lernleistung. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie: Bd. 2. Psychologie des Lernens und der Instruktion* (S. 119–152). Göttingen: Hogrefe.
- Spearman, C. (1904). “General Intelligence” objectively determined and measured. *American Journal of Psychology*, 15, 201–293.
- Spearman, C. (1927). *The abilities of man*. New York: Macmillan.
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für pädagogische Psychologie*, 19, 85–95.
- Stanat, P., & Kunter, M. (2001). Geschlechterunterschiede in Basiskompetenzen. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 251–270). Opladen: Leske + Budrich.
- Stanat, P., & Lüdtke, O. (in Druck). Internationale Schulleistungsvergleiche. In G. Trommsdorff & H.-J. Kornadt (Hrsg.), *Enzyklopädie der Psychologie: Bd. 2. Kulturelle Determinanten des Erlebens und Verhaltens*. Göttingen: Hogrefe.
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25 (2), 173–180.
- Stern, E. (1997). Mathematik. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie: Bd. 3. Psychologie des Unterrichts und der Schule* (S. 397–426). Göttingen: Hogrefe.
- Stern, E. (2001a). Intelligence, prior knowledge, and learning. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (pp. 7670–7674). Amsterdam: Elsevier.
- Stern, E. (2001b). Intelligenz, Wissen, Transfer und der Umgang mit Zeichensystemen. In E. Stern & J. Guthke (Hrsg.), *Perspektiven der Intelligenzforschung* (S. 163–204). Lengerich: Pabst.
- Sternberg, R. J. (2003). Contemporary theories of intelligence. In W. M. Reynolds & G. E. Miller (Eds.), *Handbook of psychology: Educational psychology* (pp. 23–46). Hoboken, NJ: Wiley.
- Sternberg, R. J., & Ben-Zeev, T. (Eds.). (1996). *The nature of mathematical thinking*. Mahwah, NJ: Erlbaum.
- Süllwold, F. (1983). Pädagogische Diagnostik. In K.-J. Groffmann & L. Michel (Hrsg.), *Enzyklopädie der Psychologie: Bd. 2. Intelligenz- und Leistungsdiagnostik* (S. 307–386). Göttingen: Hogrefe.
- Süß, H.-M. (1996). *Intelligenz, Wissen und Problemlösen*. Göttingen: Hogrefe.
- Süß, H.-M. (2001). Prädiktive Validität der Intelligenz im schulischen und außerschulischen Bereich. In E. Stern & J. Guthke (Hrsg.), *Perspektiven der Intelligenzforschung* (S. 109–135). Lengerich: Pabst.
- Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability – and a little bit more. *Intelligence*, 30, 261–288.
- Swanson, H. L., & Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology*, 96, 471–491.
- Sweller, J. (1989). Cognitive technology: Some procedures for facilitating learning and problem solving in mathematics and science. *Journal of Educational Psychology*, 81 (4), 457–466.
- Tent, L. (2001). Zensuren. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (2. Aufl., S. 805–811). Weinheim: Psychologie Verlags Union.
- Terhart, E. (2003). Die Lehrerbildung. In K. S. Cortina, J. Baumert, A. Leschinsky, K. U. Mayer & L. Trommer (Hrsg.), *Das Bildungswesen in der Bundesrepublik Deutschland. Strukturen und Entwicklungen im Überblick* (S. 787–810). Reinbek: Rowohlt.
- Thurstone, L. L. (1936). The factorial isolation of primary abilities. *Psychometrika*, 1, 175–182.
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.
- Tirre, W. C., & Pena, C. M. (1993). Components of quantitative reasoning: General and group ability factors. *Intelligence*, 17, 501–521.
- Trautwein, U., & Köller, O. (2003). The relationship between homework and achievement – still much of a mystery. *Educational Psychology Review*, 15 (2), 115–145.
- Treumann, K. (1974). *Dimensionen der Schulleistung. Leistungsdimensionen im Mathematikunterricht*. Stuttgart: Klett.



- Tryon, R. C. (1935). A theory of psychological components – an alternative to mathematical factors. *Psychological Review*, 42, 425–454.
- Turkheimer, E. (1994). Socioeconomic status and intelligence. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 992–1000). New York: Macmillan.
- Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*, 39, 111–133.
- Vernon, P. E. (1964). *The structure of human abilities* (2nd ed.). New York: Wiley.
- Vernon, P. E. (1965). Ability factors and environmental influences. *American Psychologist*, 20, 723–733.
- Vernon, P. E. (1969). *Intelligence and cultural environment*. London: Methuen & Co.
- Walberg, H. J. (1986). Syntheses of research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 214–229). New York: Macmillan.
- 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.
- Wang, W.-C., Chen, P.-H., & Cheng, Y.-Y. (2004). Improving measurement precision of test batteries using multi-dimensional item response models. *Psychological Methods*, 9, 116–136.
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54, 427–450.
- Wechsler, D. (1974). *Wechsler intelligence scale for children, revised*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (1981). *Wechsler adult intelligence scale, revised*. San Antonio, TX: Psychological Corporation.
- Weinert, F. E. (1994). Entwicklung und Sozialisation der Intelligenz, der Kreativität und des Wissens. In K. A. Schneewind (Hrsg.), *Enzyklopädie der Psychologie: Bd. 1. Psychologie der Erziehung und Sozialisation* (S. 259–284). Göttingen: Hogrefe.
- Weinert, F. E. (1996). Lerntheorien und Instruktionsmodelle. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie: Bd. 2. Psychologie des Lernens und der Instruktion* (S. 1–48). Göttingen: Hogrefe.
- Weinert, F. E. (2001). Schulleistungen – Leistungen der Schule oder der Schüler? In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 73–86). Weinheim: Beltz.
- Weinert, F. E., & Helmke, A. (1995a). Interclassroom differences in instructional quality and interindividual differences in cognitive development. *Educational Psychologist*, 30 (1), 15–20.
- Weinert, F. E., & Helmke, A. (1995b). Learning from wise mother nature or big brother instructor: The wrong choice as seen from an educational perspective. *Educational Psychologist*, 30 (3), 135–142.
- Weinert, F. E., & Helmke, A. (1998). The neglected role of individual differences in theoretical models of cognitive development. *Learning and Instruction*, 8 (4), 309–323.
- White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, 91, 461–481.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25, 68–81.
- Wilhelm, O. (2000). *Psychologie des schlussfolgernden Denkens. Differentialpsychologische Prüfung von Strukturüberlegungen*. Hamburg: Kovac.
- Wilhelm, O. (2005). Measuring reasoning ability. In O. Wilhelm & R. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 373–392). Mahwah, NJ: Erlbaum.
- Wittmann, W. W. (1988). Multivariate reliability theory: Principles of symmetry and successful validation strategies. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 505–560). New York: Plenum.
- Wittmann, W. W. (2005). Group differences in intelligence and related measures. In O. Wilhelm & R. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 223–240). Mahwah, NJ: Erlbaum.
- Wu, M. L. (1997). *The development and application of a fit test for use with marginal maximum likelihood estimation and generalised item response models*. University of Melbourne.
- Wu, M. L., Adams, R. J., & Wilson, M. R. (1998). *ACER Conquest*. Melbourne: The Australian Council for Educational Research Ltd.
- Zieky, M. (1994). Achievement testing. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 688–691). New York: Macmillan.
- Zielinski, W. (1996). Lernschwierigkeiten. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie: Bd. 2. Psychologie des Lernens und der Instruktion* (S. 369–402). Göttingen: Hogrefe.
- Zimmer, K., Burba, D., & Rost, J. (2004). Kompetenzen von Jungen und Mädchen. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 211–223). Münster: Waxmann.

## **Selbstständigkeitserklärung**

Ich erkläre hiermit, dass die vorliegende Dissertation ohne fremde Hilfe selbst verfasst wurde und nur die angegebenen Quellen oder Hilfsmittel benutzt wurden. Wörtliche oder sinngemäß aus anderen Quellen übernommene Stellen sind unter Angabe der Quellen kenntlich gemacht.