J Math Didakt
DOI 10.1007/s13138-011-0026-2

ORIGINALARBEIT

Eine alternative Modellierung von Geschlechtsunterschieden in Mathematik

Martin Brunner · Stefan Krauss · Laura Martignon

Eingegangen: 14. September 2010 / Angenommen: 16. Juni 2011

Zusammenfassung Geschlechtsunterschiede in Mathematik können mittels alternativer Messmodelle analysiert werden. Im Standardmodell der empirischen Bildungsforschung wird angenommen, dass die Leistung bei Mathematikaufgaben nur von einer mathematischen Kompetenz und nicht von der Intelligenz der Schülerinnen und Schüler funktional abhängig ist. Das Nested-Faktormodell hingegen basiert auf zentralen theoretischen Annahmen der Intelligenzforschung und geht davon aus, dass die Leistung bei Mathematikaufgaben auch von Intelligenz abhängt. Üblicherweise ergeben sich bei Verwendung des Standardmodells geringe Geschlechtsunterschiede zu Gunsten der Jungen. Im vorliegenden Artikel zeigen wir auf Grundlage der Daten von Jugendlichen in der 9. Klasse ($N = 29.171$), die an der deutschen PISA-2000 Studie teilnahmen, dass mit einer Modellierung mathematischer Kompetenz mittels des Nested-Faktormodells deutlich größere Geschlechtsunterschiede zu Gunsten der Jungen resultieren. Wir diskutieren diese Befunde hinsichtlich einer Neubewertung von Geschlechtsunterschieden in Mathematik.

Schlüsselwörter Geschlechtsunterschiede · Mathematik · PISA · Konfirmatorische Faktorenanalyse

M. Brunner (✉)
EMACS Forschungsgruppe, Universität Luxemburg, Campus Walferdange, 7201 Walferdange, Luxembourg
e-mail: martin.brunner@uni.lu

S. Krauss
Fakultät für Mathematik, Didaktik der Mathematik, Universität Regensburg, Universitätsstr. 31, 93053 Regensburg, Deutschland
e-mail: stefan1.krauss@mathematik.uni-r.de

L. Martignon
Institut für Mathematik und Informatik, Pädagogische Hochschule Ludwigsburg, Reuteallee 46, 71634 Ludwigsburg, Deutschland
e-mail: martignon@ph-ludwigsburg.de

Published online: 06 August 2011

Mathematics Subject Classification (2000) 97C20 · 97C40 · 97C99

Alternative Measurement Models to Assess Gender Differences in Mathematics

Abstract Gender differences in mathematics can be assessed by means of different, alternative measurement models. In the Standard Model of empirical education research it is usually assumed that performance in solving mathematical tasks depends only upon a mathematical competency. By contrast the Nested-Factor Model, based on theoretical tenets of intelligence research, assumes that performance in solving mathematical tasks is also dependent upon mere intelligence. Usually the gender differences assessed when using the Standard Model are small and favor the boys. In this paper we analyze the German data of ninth graders ($N = 29.171$), who participated in the PISA-2000 Study, by modeling mathematical competency based on the Nested-Factor Model; this analysis results in much larger gender differences in favor of boys. We discuss these findings in the light of a new evaluation of gender differences in mathematics.

1 Einleitung

Geschlechtsunterschiede in Mathematik haben große Relevanz für Forschung, pädagogische Praxis und politische Entscheidungen. So ist auch der Ausgleich jeglicher Geschlechtsunterschiede in Mathematik ein erklärtes pädagogisches Ziel der Mathematik-Didaktik (Budde 2009; Heinze et al. 2007; Leder und Forgasz 2008; Martignon et al. 2006; Zentralblatt für Didaktik der Mathematik 1994).

Zahlreiche Studien fanden bislang, dass Jungen in Mathematik geringfügig besser sind als Mädchen (Hedges und Nowell 1995; Hyde 2005; Hyde et al. 1990; Organisation for Economic Co-operation and Development 2004, 2007, 2001; Stanat und Kunter 2001; Zimmer et al. 2004). Im vorliegenden Artikel zeigen wir, dass die Größe von Geschlechtsunterschieden in Mathematik stark vom verwendeten Messmodell mathematischer Kompetenz abhängt und dass bei einer alternativen Modellierung (mit dem sog. „Nested-Faktormodell") sogar noch deutlich größere Geschlechtsunterschiede zu Gunsten der Jungen resultieren. Hierzu erläutern wir detailliert messtheoretische Grundlagen zur Erfassung mathematischer Kompetenz und der Bestimmung von Geschlechtsunterschieden in Mathematik. Wir illustrieren dann mit Rückgriff auf die Ergebnisse der Studie von Brunner et al. (2008a) unsere theoretischen Überlegungen empirisch anhand der Daten von über 29.000 deutschen Schülerinnen und Schülern in der 9. Klasse, die an der PISA-2000 Studie (Baumert et al. 2002) teilnahmen. Weiterhin berichten wir auf Grundlage dieser Datenbasis neue Befunde, welche die Prozentrechnung als Bereich mit besonders großen Geschlechtsunterschieden identifizieren. Abschließend synthetisieren wir die Ergebnisse aus mehreren empirischen Studien (Brunner 2005, 2008; Brunner und Krauss 2010; Brunner et al. 2008b), die den Zusammenhang der mathematischen Kompetenz mit sprachlichen und mathematischen Schülermerkmalen mit den Daten aus PISA 2000 untersuchen. Hierbei integrieren wir auch neue Ergebnisse zu Geschlechtsunterschieden in diesen Schülermerkmalen. Unsere Befunde motivieren eine Diskussion der

Bedeutung der mathematischen Kompetenz für schulische Entwicklungsprozesse und die Berufswahl und geben mögliche Ansatzpunkte für einen Ausgleich von Geschlechtsunterschieden in Mathematik.

2 Forschungsstand zur Größe von Geschlechtsunterschieden in mathematischer Kompetenz

Dieser Abschnitt gibt einen Überblick zur bisherigen Forschung zur Größe von Geschlechtsunterschieden in Mathematik. Schwerpunkt bilden dabei die Ergebnisse repräsentativer nationaler und internationaler quantitativer Studien für die in diesem Beitrag fokussierte Sekundarstufe. Das Effektstärken-Maß d (Cohen 1992) ermöglicht es, studienübergreifend Resultate zusammenfassen, auch wenn nicht der gleiche Test verwendet wurde: d wird berechnet, indem vom Mittelwert der Jungen der Mittelwert der Mädchen abgezogen wird. Diese Differenz wird dann durch die gemeinsame Standardabweichung geteilt. Positive Werte von d indizieren einen Leistungsvorsprung der Jungen, negative Werte einen Leistungsvorsprung der Mädchen.[1] In der pädagogisch-psychologischen Forschung werden absolute Werte von d folgendermaßen interpretiert (s. hierzu Hyde 2005): $0 < |d| \leq .10$ als „nahe bei Null", $.10 < |d| \leq .35$ als „klein", $.35 < |d| \leq .65$ als „mittel", $.66 < |d| \leq 1.00$ als „groß" und $|d| > 1.00$ als „sehr groß".

Tabelle 1 fasst bedeutende large-scale Studien und Meta-Analysen zusammen, die Geschlechtsunterschiede in der mathematischen Kompetenz in der Sekundarstufe analysieren (für eine Zusammenfassung der Ergebnisse zu internationalen Vergleichsstudien, s. a. Exekutivagentur Bildung (EACEA P9 Eurydice) 2010, bzw. Organisation for Economic Co-operation and Development 2009).

Die wichtigsten und umfangreichsten Meta-Analysen zu Geschlechtsunterschieden in Mathematik stammen aus der Arbeitsgruppe von Janet Hyde. Hyde et al. (1990) untersuchten 100 Einzelstudien (vornehmlich aus dem anglo-amerikanischen Raum) mit Daten von insgesamt über einer Million Kindern, Jugendlichen und Erwachsenen im Rahmen einer Meta-Analyse. Dies stellt die umfangreichste quantitative Überblicksarbeit zu diesem Thema für Studien bis zum Jahr 1990 dar. Ein zentrales Ergebnis dieser Meta-Analyse in Bezug auf die Sekundarstufe war, dass Jungen in der Altersgruppe von 15 bis 18 Jahren gegenüber Mädchen einen kleinen Leistungsvorsprung in Mathematik hatten (die mittlere Effektstärke über 53 Studien in dieser Altersgruppe lag bei $d = .29$). In der aktuellsten Meta-Analyse analysierten Lindberg und Kolleginnen (Lindberg et al. 2010) für die gleichen Personenpopulationen (mit Daten von über 1,2 Million Personen) die Geschlechtsunterschiede in Mathematik aus 242 Studien (für den Zeitraum von 1990 bis 2007; wiederum vornehmlich aus dem anglo-amerikanischen Raum). Auf Grundlage der Daten von 110 Studien zu Geschlechtsunterschieden in der Sekundarstufe zeigte sich erneut, dass Jungen einen leichten Leistungsvorsprung hatten (die mittlere Effektstärke über alle

[1] In diesem Fall ist der weitverbreitete Begriff „Effektstärke" etwas unglücklich, da keinerlei Kausalannahmen zum Zusammenhang von Geschlecht und Leistungsdifferenzen gemacht werden, sondern nur die Größe des Unterschieds interessiert.

Tab. 1 Geschlechtsunterschiede in der mathematischen Kompetenz bei Jugendlichen in der Sekundarstufe

Studie	d	Alter/Klassenstufe
Meta-Analysen von Hyde und Kolleginnen		
Hyde et al. (1990): mittlere Effektgröße über insgesamt 53 Einzelstudien (Zeitraum: bis 1990)	+.29	15–18 Jahre
Lindberg et al. (2010): mittlere Effektgröße über insgesamt 110 Einzelstudien (Zeitraum: von 1990 bis 2007)	+.23	High School (14–18 Jahre)
Hedges und Nowell (1995): *US-Schulleistungsstudien*		
Project Talent (1960)	+.12	15 Jahre
National Longitudinal Study (1972)	+.24	12. Klasse
National Longitudinal Study of Youth (1980)	+.26	15–22 Jahre
High School and Beyond (1980)	+.22	12. Klasse
National Educational Longitudinal Study (1988)	+.03	8. Klasse
TIMSS (Trends in the International Mathematics and Science Study)[a]		
1995: gemittelt über alle Teilnehmerstaaten	+.08	8. Klasse
1995: nur Deutschland	+.03	8. Klasse
1999: gemittelt über alle Teilnehmerstaaten	+.04	8. Klasse
2003: gemittelt über alle Teilnehmerstaaten	+.01	8. Klasse
2007: gemittelt über alle Teilnehmerstaaten	.00	8. Klasse
PISA (Programme for International Student Assessment)		
2000: gemittelt über alle OECD-Teilnehmerstaaten	+.11	15 Jahre
2000: nur Deutschland	+.15	15 Jahre
2003: gemittelt über alle OECD-Teilnehmerstaaten	+.11	15 Jahre
2003: nur Deutschland	+.09	15 Jahre
2006: gemittelt über alle OECD-Teilnehmerstaaten	+.11	15 Jahre
2006: nur Deutschland	+.20	15 Jahre
2009: gemittelt über alle OECD-Teilnehmerstaaten	+.12	15 Jahre
2009: nur Deutschland	+.16	15 Jahre

Anmerkung. Positive Werte der Effektgröße d bedeuten, dass Jungen bessere Leistungen in Mathematik erzielten als Mädchen

[a]Seit 1995 beteiligt sich Deutschland nicht mehr an TIMSS in der Sekundarstufe. Die Anzahl der Teilnehmerstaaten variierte zwischen den Erhebungszyklen von TIMSS, was die zyklenübergreifende Vergleichbarkeit der Geschlechtsunterschiede einschränkt

Studien lag bei $d = .23$). Darüber hinaus zeigten beide Meta-Analysen von Hyde und Kolleginnen, dass in der Primarstufe Jungen und Mädchen nahezu gleich gut sind, die Leistungsschere zu Gunsten der Jungen geht also erst in der Sekundarstufe auf (vgl. Beller und Gafni 1996, Tab. 2).

Hedges und Nowell (1995) reanalysierten für ihren in *Science* publizierten Artikel fünf repräsentative US-Schulleistungsstudien. Sie fanden ebenfalls (mit einer Ausnahme, siehe Tab. 1) einen kleinen Leistungsunterschied in Mathematik zu Gunsten der Jungen.

Auch in allen vier bisherigen Erhebungszyklen von PISA wurden in nahezu allen Teilnehmerstaaten bei 15-jährigen Jugendlichen kleine Leistungsunterschiede zu Gunsten der Jungen in Mathematik festgestellt (Organisation for Economic Co-operation and Development 2001, 2004, 2007, 2010). Tabelle 1 verdeutlicht, dass dies sowohl für die Leistungsdifferenz zwischen Jungen und Mädchen gemittelt über alle OECD-Teilnehmerstaaten als auch für Deutschland gilt.

Ein etwas anderes Bild zeigt sich für die vier Erhebungszyklen von TIMSS, an denen repräsentative Stichproben von Schülerinnen und Schüler in der 4. und 8. Klasse teilnahmen (Beaton et al. 1996; Mullis et al. 2000, 2004, 2008). In Tab. 1 sind die Ergebnisse für die 8. Klasse dargestellt, da diese für die vorliegende Arbeit primär relevant sind. In allen vier Erhebungszyklen von TIMSS zeigte sich, dass sich in der Sekundarstufe Jungen und Mädchen nahezu nicht in ihrer mathematischen Kompetenz unterscheiden (detaillierte Analysen zu Geschlechtsunterschieden auf Grundlage der Daten von TIMSS 1995 finden sich auch in Kaiser und Steisel 2000).

Wie ist der Unterschied zwischen den TIMSS-Studien und den anderen Studien, die in Tab. 1 aufgelistet sind, zu erklären? Wir haben bislang Unterschiede in der generellen mathematischen Kompetenz betrachtet. Eine interessante Frage ist nun, in welchen Teilbereichen (z.B. Algebra, Geometrie, usw.) beziehungsweise bei welchen mathematischen Teilkompetenzen (z.B. Argumentieren, Modellieren, Problemlösen, usw., s. z.B. Blum et al. 2006) Geschlechtsunterschiede verstärkt auftreten. Tendenziell finden sich – sowohl international als auch in deutschen Untersuchungen – vor allem in Geometrie und Analysis Geschlechtsunterschiede (und weniger in Arithmetik und Algebra, z.B. Budde 2009; Hyde et al. 1990; Organisation for Economic Co-operation and Development 2009; Kaiser und Steisel 2000). Weiterhin zeigen sich Geschlechtsunterschiede zu Gunsten der Jungen im Besonderen bei komplexeren Problemstellungen, vor allem, wenn unkonventionelle Lösungsstrategien erforderlich sind (Budde 2009; Gallagher et al. 2000; Schwank et al. 2003; s. aber auch Kaiser und Steisel 2000). Interessanterweise zeigen sich dagegen bei kalkülorientierten Standardaufgaben keine Geschlechtsunterschiede oder es erzielen Mädchen sogar geringfügig bessere Leistungen (Budde 2009; Hyde et al. 1990, 2008). Diese differenziellen Befundmuster zu Geschlechtsunterschieden in mathematischen Teilbereichen liefern damit eine mögliche Erklärung für die im Vergleich zu anderen Studien geringeren Geschlechtsunterschiede in Mathematik bei TIMSS: So fokussiert beispielsweise PISA deutlich stärker auf Modellierungs- und Problemlöseaufgaben, wohingegen der TIMSS-Mathematiktest kalkülorientierte Fertigkeiten stark betont (Else-Quest et al. 2010, S. 105).

Insgesamt kristallisiert sich im Hinblick auf die referierten Befunde ein relativ robustes Befundmuster zu Geschlechtsunterschieden in Mathematik heraus: (1) Geschlechtsunterschiede in der (generellen) mathematischen Kompetenz sind klein (d's $< .35$). Mit Blick auf Teilbereiche, bzw. -kompetenzen sind Jungen in der Sekundarstufe (2) bei Aufgaben aus dem Bereich Geometrie und Analysis und (3) bei Problemlöse- und Modellierungsaufgaben geringfügig besser als Mädchen (Hyde et al. 1990; Lindberg et al. 2010; Organisation for Economic Co-operation and Development 2009).

Im Kontext dieses Fazits möchten wir jedoch auf einen wichtigen Punkt hinweisen. Die hier referierten Befunde, die insgesamt ein relativ robustes Muster zu Geschlechtsunterschieden in Mathematik zeigen, geben einen zusammenfassenden Überblick über zentrale Tendenzen zu diesem Forschungsthema. Dies schließt damit aber natürlich nicht aus, dass Einzelstudien, beziehungsweise einzelne Länder vom generellen Trend abweichen. So erzielten beispielsweise in der Studie von Winkelmann et al. (2008) zur Evaluation der deutschen Bildungsstandards Jungen *bereits in der 3. beziehungsweise 4. Klasse* einen Leistungsvorsprung von $d = .34$, beziehungsweise $d = .27$. Ebenso variieren die Geschlechtsunterschiede in Mathematik zum Teil erheblich zwischen verschiedenen Nationen (s.a. Else-Quest et al. 2010): So lagen beispielsweise bei PISA 2009 (Organisation for Economic Co-operation and Development 2010) die Geschlechtsunterschiede zwischen $d = -.11$ (in Albanien zu Gunsten der Mädchen) und $d = +.32$ (in Kolumbien zu Gunsten der Jungen). Darüber hinaus deutet zum Beispiel die Studie von Beller und Gafni (1996, Tab. 2) auch daraufhin, dass sich in einigen Ländern (z.B. Korea) der Entwicklungstrend umkehrt und sich die Leistungsschere mit zunehmenden Alter schließt.

Eine detaillierte Übersicht zu aktuellen Forschungsergebnissen und Theorien zur Erklärung von Geschlechtsunterschieden in Mathematik findet sich bei Budde (2009), Else-Quest und Kolleginnen, Gallagher und Kaufman (2005) sowie in Leder und Forgasz (2008). In diesen Arbeiten werden nicht nur Befunde quantitativer und qualitativer Studien zu Leistungsunterschieden, sondern auch zu affektiven und motivationalen Geschlechtsunterschieden in Bezug auf das Unterrichtsfach Mathematik diskutiert.

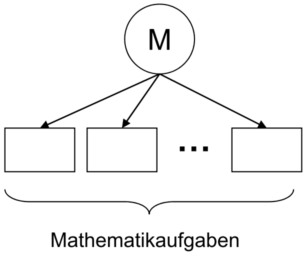
3 Messmodelle mathematischer Kompetenz

Worin *genau* unterscheiden sich nun die Schülerinnen von den Schülern? Mit anderen Worten: Was verbirgt sich bei den Studien aus Tab. 1 hinter dem Konstrukt „mathematische Kompetenz“?

An dieser Stelle ist es erstens wichtig festzuhalten, dass mathematische Kompetenz unterschiedlich theoretisch konzeptualisiert werden kann, und wie die Diskussion von TIMSS und PISA zeigte, unterschiedliche Teilbereiche der Mathematik unterschiedlich stark gewichtet in das Konstrukt „mathematische Kompetenz“ eingehen können. Zweitens, und für diesen Artikel entscheidend, ist es wichtig festzustellen, dass alle bisher referierten Befunde auf der gleichen psychometrischen Modellierung mathematischer Kompetenz basieren. In diesem Modell ist mathematische Kompetenz M ein theoretisches („latentes“, also nicht direkt beobachtbares) Konstrukt, das in Abb. 1 symbolisch durch einen Kreis dargestellt ist. Die Pfeile von M auf die Testaufgaben („manifeste“, also direkt messbare Variablen werden durch Rechtecke dargestellt) bedeuten, dass zur Lösung von Mathematik-Aufgaben die Kompetenz M benötigt wird. Die kausale Annahme ist dabei: Je höher M ausgeprägt ist, desto mehr Punkte erzielt man voraussichtlich bei den Mathematik-Testaufgaben.

In den größeren Studien der empirischen Bildungsforschung (z.B. bei PISA) wird oft auch die Intelligenz (die wir nachfolgend mit IQ abkürzen) der teilnehmenden

Abb. 1 Typisches Messmodell mathematischer Kompetenz M. Das latente Konstrukt M wird durch mehrere manifeste Mathematikaufgaben gemessen



Schülerinnen und Schüler erhoben, wobei IQ durch Intelligenztestaufgaben gemessen wird.[2] Dabei werden üblicherweise beide Kompetenzen (M und IQ) getrennt erhoben und dargestellt. Abbildung 2a zeigt ein entsprechendes Modell. Für IQ gilt wie für M: Je höher IQ ausgeprägt ist, desto mehr Punkte erzielt man bei den Intelligenztestaufgaben. Das Modell aus Abb. 2a liegt (implizit oder explizit) allen Schlussfolgerungen zu Geschlechtsunterschieden zugrunde. So nutzten beispielsweise alle in den Meta-Analysen von Hyde und Kolleginnen untersuchten Studien Varianten dieses Messmodells der mathematischen Kompetenz und auch in PISA oder TIMSS werden im Kontext von Item-Response Modellen Varianten dieses Messmodells verwendet. Wir nennen es deshalb im Folgenden „Standardmodell“, welches das „klassische“ (Bollen und Lennox 1991, S. 312) und am meisten genutzte Messmodell in der pädagogischen und psychologischen Forschung ist (vgl. Jarvis et al. 2003).

Entscheidend für unsere folgende Argumentation ist nun, dass im Standardmodell theoretisch angenommen wird, dass Intelligenz *nicht* zur Lösung von Mathematikaufgaben erforderlich ist (es gehen keine Pfeile von IQ auf die Mathematikaufgaben). Es wird zwar angenommen, dass Intelligenz und mathematische Kompetenz *auf Konstruktebene* zusammenhängen (der Doppelpfeil zwischen den beiden Konstrukten repräsentiert deren Korrelation); die Korrelation zwischen M und IQ, die man in PISA 2000 zwischen diesen beiden Konstrukten fand, ist in Abb. 2g in Form eines Streudiagramms dargestellt. Wichtig ist aber festzuhalten, dass im Standardmodell die implizite Annahme ist, dass die zur Messung der beiden Konstrukte eingesetzten Testaufgaben jeweils nur mit einer der beiden Kompetenzen gelöst werden (vgl. Modellgleichung in Abb. 2c).[3]

Das Standardmodell dominiert in der *pädagogischen und psychologischen Forschung* (Bollen und Lennox 1991; Jarvis et al. 2003). Im Gegensatz hierzu wird derzeit in der *Intelligenzforschung* (einem Teilgebiet der Psychologie) die Beziehung zwischen Intelligenz und mathematischer Kompetenz grundlegend anders aufgefasst. Insbesondere zeigen seit Spearman’s (1904) wegweisender Studie hunderte von Untersuchungen, dass kognitive Leistungen über fachspezifische Domänengrenzen hinweg positiv korrelieren (Carroll 1993; Gustafsson und Undheim 1996; Jensen 1998). Diese positiven Korrelationen können durch eine *generelle kognitive Fähigkeit* erklärt werden, die wir nachfolgend zur besseren Unterscheidung der

[2] Für eine ausführliche Diskussion des Konstrukts Intelligenz und seiner Messung siehe z.B. Hunt und Carlson (2007) oder Neisser et al. (1996).

[3] In den „Rest“ gehen z.B. Messfehler ein, die wir der Übersichtlichkeit halber nicht in die grafischen Darstellungen des Standard- und Nested-Faktormodell mit aufgenommen haben.

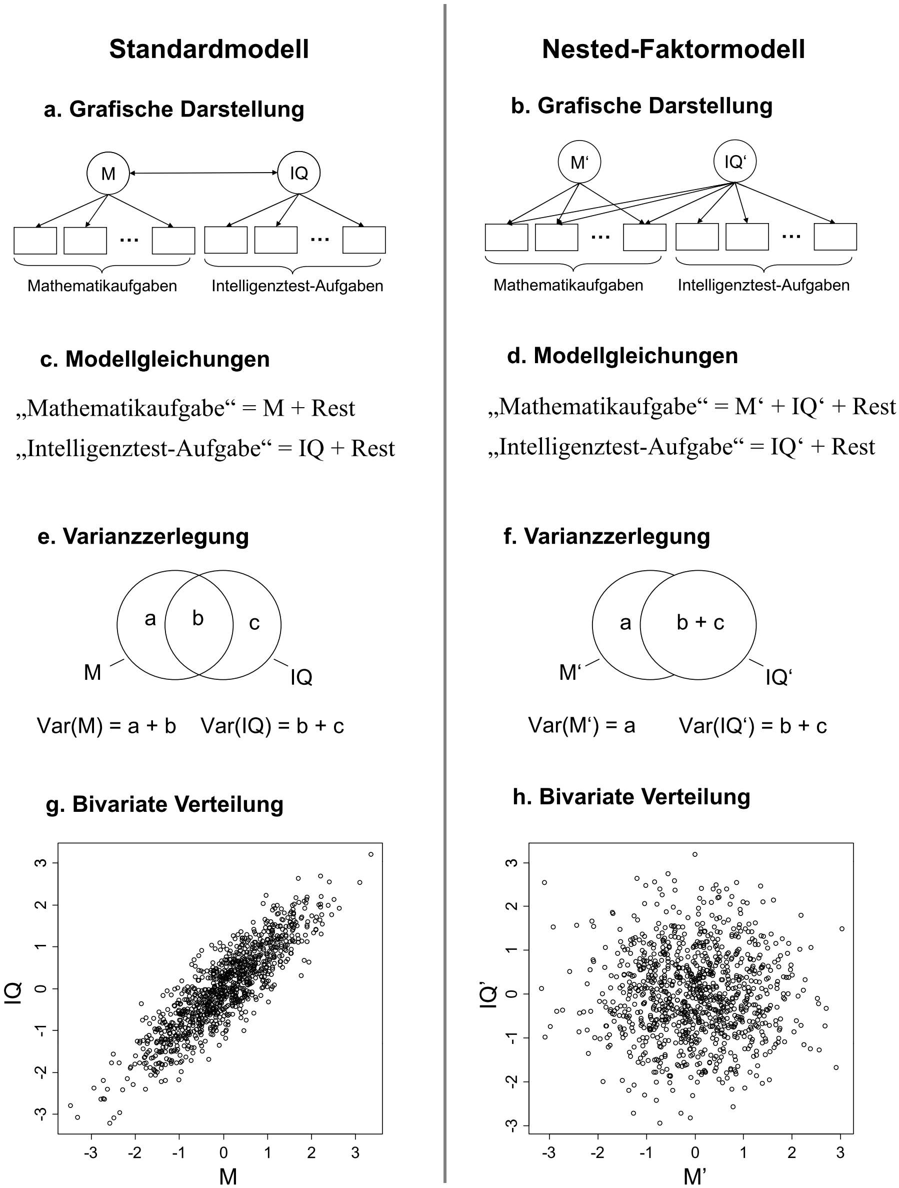


Abb. 2 Standardmodell vs. Nested-Faktormodell (**a**), (**b**): Grafische Darstellung; (**c**), (**d**): Modellgleichung; (**e**), (**f**): Varianzzerlegung; (**g**), (**h**): Illustration der bivariaten Verteilung mit jeweils standardnormalverteilten Variablen [$M = 0$, $SD = 1$] und einer Korrelation von $r = .89$ [dies entspricht der Korrelation in der bei PISA 2000 untersuchten Stichprobe, s. Brunner (2008)] im Standardmodell, bzw. $r = 0$ im Nested-Faktormodell)

beiden Modelle IQ' nennen (in der Intelligenzforschung wird diese Fähigkeit üblicherweise mit g für „general factor“, bzw. „general cognitive ability“ bezeichnet). Im Rahmen der Intelligenzforschung gehen daher aktuelle Messmodelle kognitiver

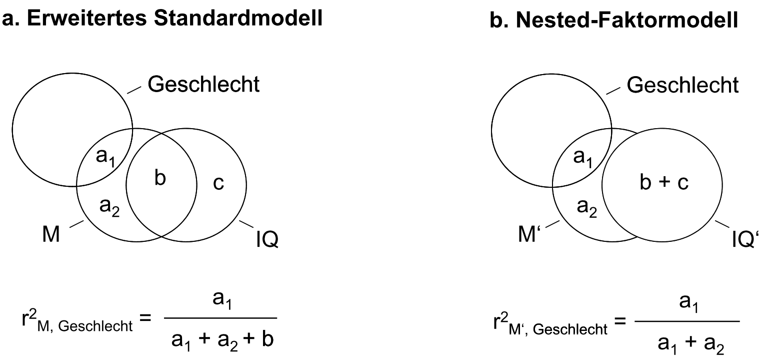


Abb. 3 Erweiterte Varianzzerlegung: Geschlechtunterschiede in Mathematik im (**a**) Standardmodell und (**b**) im Nested-Faktormodell

Kompetenzen, wie das Nested-Faktormodell (Abb. 2b, Gustafsson und Balke 1993), davon aus, dass die Leistung bei Mathematikaufgaben einerseits von einer mathematikspezifischen Kompetenz M' aber andererseits eben auch von IQ' beeinflusst wird. Im Nested-Faktormodell können bessere Leistungen bei Mathematikaufgaben also durch eine höhere Ausprägung von M' und/oder eine höhere Ausprägung von IQ' erzielt werden (vgl. Modellgleichung in Abb. 2d).

Offensichtlich unterscheiden sich Schülerinnen und Schüler sowohl in ihrer mathematischen Kompetenz als auch in ihrer Intelligenz. Das Standardmodell und das Nested-Faktormodell unterscheiden sich in der *Erklärung* dieser interindividuellen Unterschiede; die beiden Modelle machen jeweils unterschiedliche Annahmen zur Varianzzusammensetzung der beiden latenten Variablen. Die Varianzzusammensetzung wird in den Abb. 2e (Standardmodell) und 2f (Nested-Faktormodell) dargestellt (die Varianz aller Variablen wurde hierbei auf 1 standardisiert; vgl. Cohen et al. 2003). Im Standardmodell korrelieren M und IQ (wie Abb. 2g illustriert). Die Fläche b repräsentiert diese gemeinsame Varianz, b entspricht dabei der quadrierten Korrelation r^2 zwischen M und IQ. Die Fläche a repräsentiert die Varianz, die M *nicht* mit IQ teilt; die Fläche c repräsentiert die Varianz, die IQ *nicht* mit M teilt. Die Gesamtvarianz von M entspricht also der Summe aus a und b; die Gesamtvarianz von IQ entspricht der Summe aus b und c.

Eine zentrale Annahme im Nested-Faktormodell ist, dass die Konstrukte M' und IQ' *auf Konstruktebene* voneinander unabhängig sind. Es besteht also kein systematischer, korrelativer Zusammenhang zwischen M' und IQ' (deshalb ist im Modell 2b kein Korrelationspfeil mehr eingezeichnet). Das Streudiagramm von M' und IQ' (Abb. 2h) illustriert die fehlende Korrelation zwischen den beiden latenten Variablen: Man sieht, dass ein Teil der Jugendlichen mit überdurchschnittlich hohen Werten auf M' (bei standardnormalverteilten Variablen sind dies Werte über Null) überdurchschnittlich hohe Werte auf IQ' und ein in etwa gleich großer Anteil unterdurchschnittlich niedrige Werte auf IQ' hat.

Es ist wichtig darauf hinzuweisen, dass sich im Vergleich zum Standardmodell die beobachtete Gesamtvarianz nicht ändert; es ändert sich jedoch im Nested-Faktormodell die Art der Zerlegung der Gesamtvarianz. Die Gesamtvarianz von M' ist in Abb. 2f durch die Fläche a repräsentiert; M' besitzt keine gemeinsame Varianz

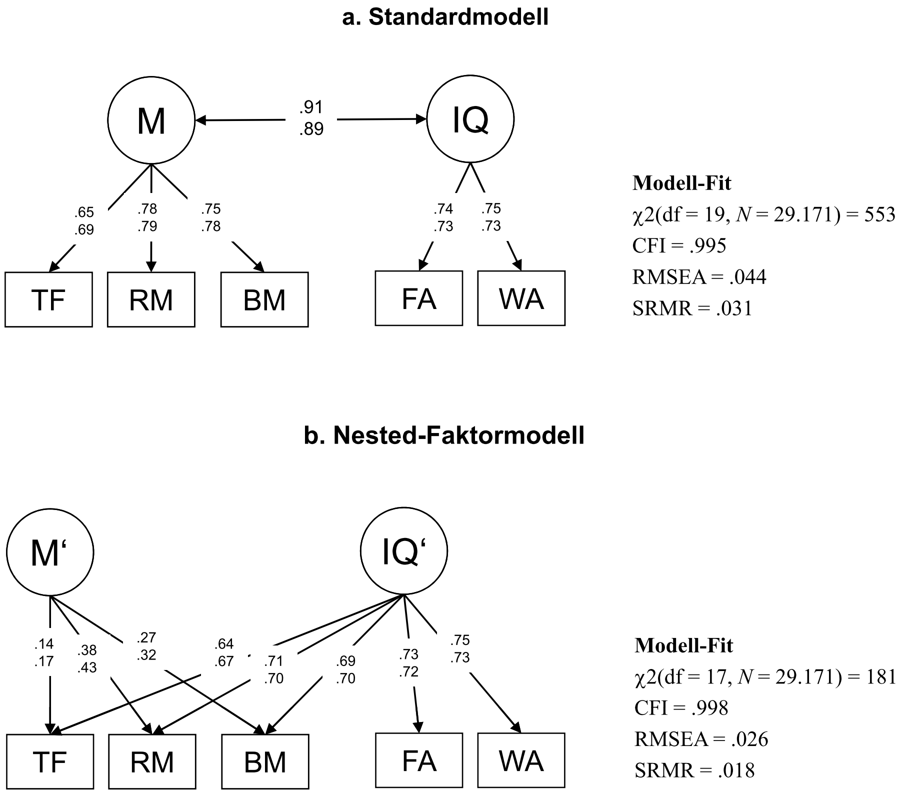


Abb. 4 Standardisierte Modellparameter (an den Pfeilen *obere Zahl*: Mädchen; *untere Zahl*: Jungen) und Modell-Fit des (**a**) erweiterten Standardmodells und (**b**) Nested-Faktormodells. Erklärung der Abkürzungen: *TF*: technische Fertigkeiten; *RM*: rechnerisches Modellieren; *BM*: begriffliches Modellieren; *FA*: Figurenanalogien; *WA*: Wortanalogien; χ^2: Goodness-of-Fit-Statistik; *df*: Anzahl der zugehörigen Freiheitsgrade; *CFI*: Comparative fit Index; *RMSEA*: Root Mean Square Error of Approximation; *SRMR*: Standardized Root Mean Squared Residual (diese Fit-Indizes werden z.B. erläutert in Hu und Bentler 1998)

mit IQ'. Da M' somit eine reine (sozusagen „intelligenzfreie“) mathematische Kompetenz repräsentiert, haben wir sie *mathematikspezifische* Kompetenz genannt. Die Varianz von IQ' wird nun durch die gemeinsame Fläche „b + c“ repräsentiert.

Was unterscheidet M von M', bzw. *IQ von* IQ'? Während IQ theoretisch lediglich Intelligenztest-Aufgaben beeinflusst, hat IQ' Einfluss auf Intelligenztest-Aufgaben *und* auf Mathematikaufgaben. Sowohl im erweiterten Standardmodell als auch im Nested-Faktormodell wird Intelligenz jedoch primär durch Intelligenztestaufgaben gemessen. Im Nested-Faktormodell wird zwar zusätzlich angenommen, dass Intelligenz *außerdem* zum Lösen von Mathematikaufgaben hilfreich ist, dies beeinflusst aber nicht die konzeptionelle Vorstellung über das Konstrukt. Der Unterschied in beiden Modellen besteht also in der angenommenen *Wirkung* des Konstrukts. Zwei Gründe sprechen für die Annahme im Nested-Faktormodell, dass IQ' nur durch Intelligenzaufgaben gemessen wird. Erstens, Gustafsson (1984, 1988) hat mehrere Studien zur Intelligenzstruktur auf Grundlage der Daten von großen, heteroge-

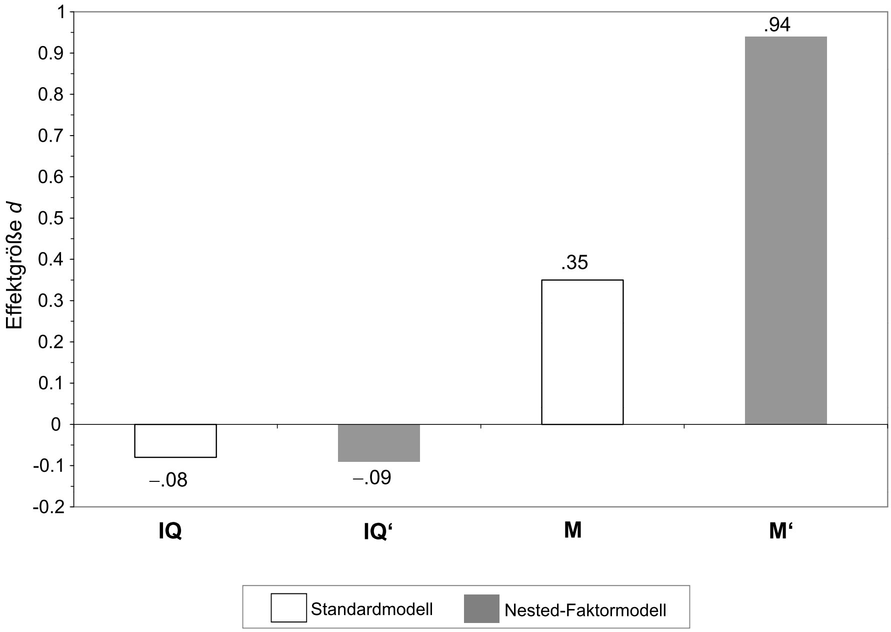


Abb. 5 Geschlechtsunterschiede in IQ, IQ', M und M'. Positive Werte der Effektgröße d bedeuten, dass Jungen bessere Leistungen erzielten als Mädchen

nen Schülerstichproben (im Alter zwischen 11 und 15 Jahren) durchgeführt. Hierbei konnte er mittels eines hierarchischen Faktormodells zeigen, dass IQ und IQ' *empirisch* äquivalent sind, denn in seinem Modell wurde die Varianz von IQ perfekt durch die Varianz von IQ' erklärt (van der Sluis et al. 2005). Da man weiterhin zeigen kann, dass der Faktor IQ' aus seinem hierarchischen Faktormodell *mathematisch* äquivalent ist mit dem Faktor IQ' in unserem Nested-Faktormodell (Yung et al. 1999), setzen wir IQ und IQ' gleich: Daher sind auch die Modellgleichungen für IQ und IQ' identisch (s. Abb. 2c und 2d). Zweitens, liegt das schlussfolgernde Denken im konzeptuellen Kern von Definitionen der Intelligenz (Gottfredson 1997; Snow und Lohman 1989) und Aufgaben zum schlussfolgernden Denken befinden sich auch in nahezu allen Intelligenztests. Dies stützt das Vorgehen, die generelle kognitive Fähigkeit IQ' ausschließlich mit Aufgaben zum schlussfolgernden Denken zu messen (vgl. Eid et al. 2003). Aus den beiden genannten Gründen sind die Modellgleichungen (Abb. 2c und 2d) für IQ und IQ' identisch.

M und M' sind zwar beide Indikatoren „mathematischer Kompetenz“, sie sind jedoch – im Gegensatz zu IQ und IQ' – auch *konzeptuell* verschieden: M entspricht der Standardvorstellung mathematischer Kompetenz, das heißt, wer eine höhere mathematische Kompetenz besitzt, ist „besser in Mathematik“. Ganz entscheidend ist nun: Eine zentrale Annahme des Standardmodells ist, dass geringe Werte in M – was das Lösen der Mathematik-Testaufgaben betrifft – *nicht* durch höhere Intelligenz ausgeglichen werden können. Man kann den Modellgleichungen (Abb. 2c) entnehmen, dass die Performanz bei den Mathematik-Testaufgaben ausschließlich auf M zurückgeführt wird.

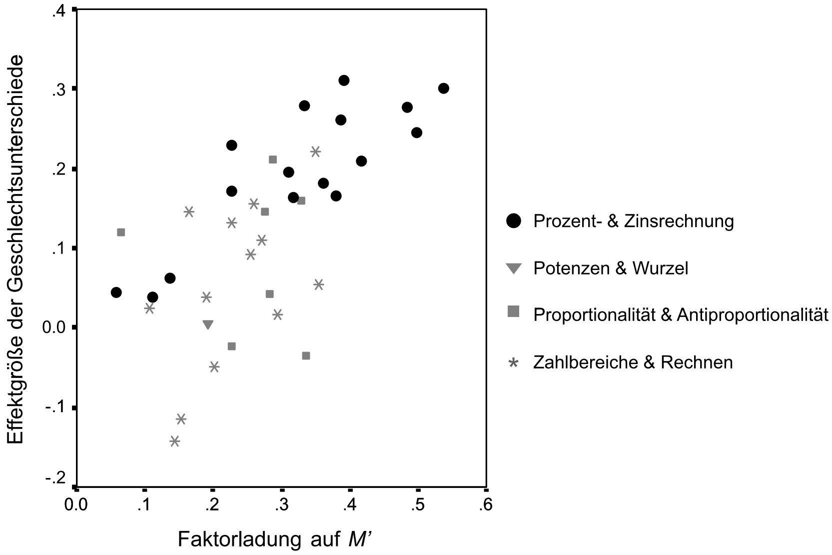


Abb. 6 Geschlechtsunterschiede und Faktorladungen für die einzelnen PISA-Arithmetik Items. Arithmetik wurde dabei aufgeteilt in die Bereiche Proportionalität und Antiproportionalität, Potenzen & Wurzeln, Prozent- und Zinsrechnung sowie Zahlbereiche und Rechnen

Hingegen wird eine Mathematikaufgabe im Nested-Faktormodell (Abb. 2b) von 2 Pfeilen „getroffen", d.h., für eine Lösung einer Mathematikaufgabe kann sowohl auf M' als auch auf IQ' „zurückgegriffen" werden: Ein Schüler kann geringe Mathematikkompetenz in diesem Modell also gegebenenfalls durch hohe Intelligenz ausgleichen.

Im Hinblick auf Geschlechtsunterschiede in „mathematischer Kompetenz" wird nun klar, dass im Standardmodell Geschlechtsunterschiede in M und im Nested-Faktormodell Geschlechtsunterschiede in M' untersucht werden. Im Folgenden werden wir sehen, dass in beiden Modellen deutlich verschiedene Geschlechtsunterschiede in „mathematischer Kompetenz" resultieren.

4 Modellierung von Geschlechtsunterschieden in Mathematik

Mit dem Standardmodell und dem Nested-Faktormodell existieren zwei alternative, theoretisch gestützte Messmodelle mathematischer Kompetenz. Jedoch wurden bislang Geschlechtsunterschiede in Mathematik im Wesentlichen mit dem Standardmodell analysiert. Ein Grund hierfür ist, dass sich Intelligenzforscher zwar unter anderem auch für Geschlechtsunterschiede interessieren, aber natürlich vornehmlich für Unterschiede in ihrem Forschungsobjekt, der Intelligenz (siehe z.B. Lynn 1999; van der Sluis et al. 2006). Diesbezüglich besteht breiter Konsens und viele Studien zeigen empirisch, dass Geschlechtsunterschiede in Intelligenz um Null liegen (Halpern und LaMay 2000; Jensen 1998; Neisser et al. 1996; siehe aber auch Lynn und Irwing 2004). Dies impliziert, dass kein systematischer Zusammenhang zwischen der

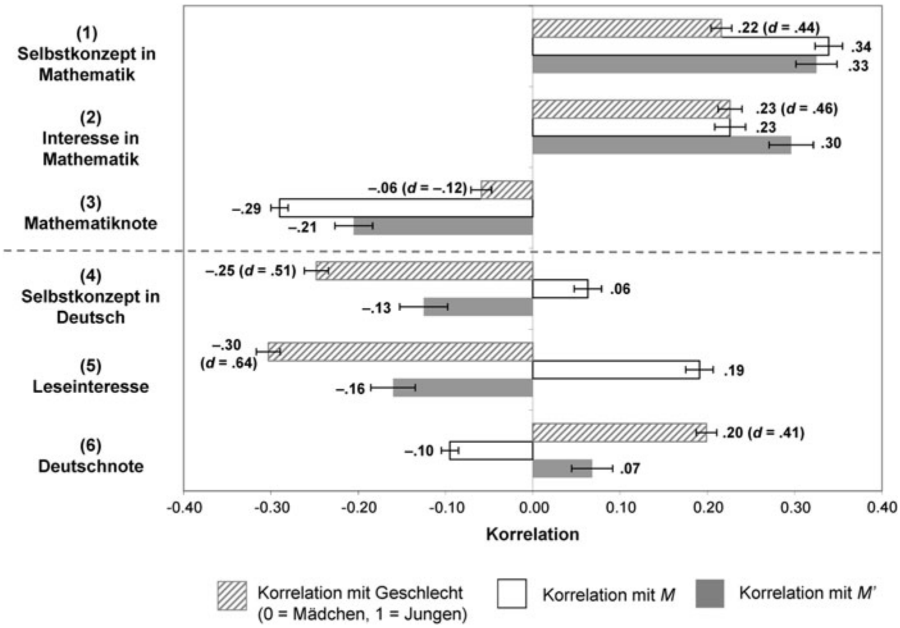


Abb. 7 Korrelationen (und zugehörige 95 %-Konfidenzintervalle) von Geschlecht, M und M' mit mathematischen und sprachlichen Schülermerkmalen. Die Korrelationen von Geschlecht mit Schülermerkmalen sind punktbiserale Korrelationen, die wir nach Cohen (1988, S. 23) in d-Werte transformiert haben

Variable Geschlecht und IQ' besteht (und somit auch nicht mit IQ). In Abb. 3a und b ist das schematisch so dargestellt, dass es keine Schnittfläche zwischen Geschlecht und den Flächen b oder c gibt.

Die bisher gefundenen geringfügigen Geschlechtsunterschiede in Mathematik zu Gunsten der Jungen sind durch die Fläche a_1, also der Schnittfläche der Geschlechtsvariablen mit M (Abb. 3a) dargestellt. Die Fläche a_2 ist der Varianzanteil, den M weder mit der Geschlechtsvariablen noch mit IQ teilt. Die Größe der Geschlechtsunterschiede in M (in diesem Fall die quadrierte Korrelation zwischen M und Geschlecht) berechnet sich somit aus dem Anteil der Fläche a_1 an der Gesamtvarianz von M (also der Summe aus a_1, a_2 und b).

Wie berechnet man Geschlechtsunterschiede in M'? Zur Erinnerung: Im Vergleich zum Standardmodell ändert sich die beobachtete Gesamtvarianz nicht, es ändert sich jedoch die Art und Weise, wie im Nested-Faktormodell diese Gesamtvarianz zerlegt wird. Die gemeinsame Varianz zwischen M' und Geschlecht ist also genauso groß wie im Standardmodell und wird durch a_1 in Abb. 3b dargestellt. Die Fläche a_2 ist der Varianzanteil, den M' weder mit der Geschlechtsvariablen noch mit IQ' gemein hat. Die Gesamtvarianz von M' ist die Summe aus a_1 und a_2 (die Varianz, die durch die Fläche b repräsentiert wird, gehört ausschließlich zu IQ'). Die Größe der Geschlechtsunterschiede in M' (in diesem Fall die quadrierte Korrelation zwischen M' und Geschlecht) berechnet sich somit aus dem Anteil der Fläche a_1 an der Gesamtvarianz von M' (also der Summe aus a_1 und a_2).

Wie verhalten sich Geschlechtsunterschiede in M zu Geschlechtsunterschieden in M'? Aus den Bruchtermen zur Berechnung der Geschlechtsunterschiede (Abb. 3a und 3b) folgt unmittelbar, dass (wenn M und IQ' korrelieren und somit $b > 0$ ist) Geschlechtsunterschiede in M' größer sein müssen als in M, da die Gesamtvarianz von M' kleiner ist als die Gesamtvarianz von M. Oder aus Sicht des Nested-Faktormodells betrachtet: Bezüglich der Intelligenz IQ' gibt es keine Geschlechtsunterschiede. Die in Mathematik üblicherweise gefundenen Geschlechtsunterschiede (also die in M) sind demnach deshalb so „gering", weil zur Lösung von Mathematikaufgaben neben der mathematikspezifischen Fähigkeit M' auch Intelligenz IQ' erforderlich ist. IQ' erklärt also auch Varianz in den Mathematikaufgaben. Diese erklärte Varianz entspricht der Fläche b im Standardmodell und geht in den Nenner zur Berechnung der Geschlechtsunterschiede in M ein, was unterschiedliche Geschlechtsunterschiede in M und M' zur Folge hat.

Die Differenz der ermittelten Geschlechtsunterschiede hängt dabei wesentlich von der Größe der gemeinsamen Varianz zwischen Intelligenztestaufgaben und Mathematikaufgaben ab: Je größer diese gemeinsame Varianz (also die Fläche b) ist, desto größer sind die Unterschiede zwischen den Geschlechtsunterschieden, die man für M beziehungsweise M' ermittelt. Wichtig ist hierbei, dass dieser gemeinsame Varianzanteil üblicherweise substantiell ist, wie beispielsweise eine Überblicksarbeit des Intelligenzforschers David Lubinski belegt: „In heterogeneous collections of cognitive tests in a wide range of talent, general intelligence accounts for roughly 50 % of the common variance [dies entspricht der Fläche b] (quantitative [dies entspricht der Summe aus a_1 und a_2], spatial, and verbal ability each account for approximately 8 %–10 % of the remaining common variance)." (Lubinski 2004, S. 98).

Das Ziel der nachfolgenden empirisch-orientierten Abschnitte ist es, (1) diese theoretische Überlegung zu Geschlechtsunterschieden in Mathematik anhand eines repräsentativen Datensatzes zu verifizieren. Wir greifen hierbei einerseits auf die Analysen von Brunner et al. (2008a) zurück und erweitern diese um detaillierte Analysen zu Geschlechtsunterschieden in mathematischen Aufgabenklassen und Stoffgebieten. Ein zentrales Anliegen ist hierbei, das Bewusstsein für alternative Möglichkeiten zur Modellierung kognitiver Kompetenzen zu wecken, insbesondere bei Forscherinnen und Forschern, die Geschlechtsunterschiede in Mathematik untersuchen. Denn das Nested-Faktormodell wurde hierzu bislang nur in sehr wenigen Studien genutzt. Eine der wenigen Studien stammt von Hartig und Höhler (2008), die Geschlechtsunterschiede in der generellen verbalen Kompetenz und spezifischen Kompetenzen für Lesen und Hören untersuchten. Weiterhin befasste sich die Studie von Rosén (1995) mit Geschlechtsunterschieden in einer Vielzahl spezifischer kognitiver Fähigkeiten und in der generellen kognitiven Fähigkeit auf Grundlage der Daten von über 1200 13-jährigen Jugendlichen aus Schweden. Rosén fand unter anderem einen mittleren Leistungsvorsprung zu Gunsten der Mädchen in der generellen kognitiven Fähigkeit ($d = -.45$) und einen sehr großen Leistungsvorsprung in der mathematikspezifischen Fähigkeit zu Gunsten der Jungen ($d = 1.32$).

5 Untersuchungsmethode

Die folgenden empirischen Analysen basieren auf den repräsentativen Daten von über 29.000 Neuntklässlerinnen und Neuntklässlern, die an der deutschen Erweiterung der PISA-2000 Studie teilnahmen (Baumert et al. 2002). Detaillierte Informationen zur Stichprobe finden sich in Brunner (2005), bzw. Brunner et al. (2008a).

Mathematische Kompetenz wurde mit den 117 Mathematikaufgaben gemessen, die in den internationalen und nationalen Tests in PISA-2000 verwendet wurden. Wenn eine Aufgabe Modellierungsprozesse erforderte, wurde sie – je nachdem ob überwiegend prozedurales oder konzeptuelles Denken erforderlich war – der Klasse der *rechnerischen* bzw. *begrifflichen Modellierungsaufgaben* zugeordnet (RM bzw. BM). Dazu kam die Klasse der *technischen Aufgaben* (TA), die nur den Abruf von Fertigkeiten oder Faktenwissen erfordern (Neubrand 2003; Neubrand et al. 2001). Für unsere Analysen wurden die erzielten Punkte bei diesen (bei der Zuordnung disjunkten und exhaustiven) Aufgabenklassen zu insgesamt drei Skalenscores für TF, RM und BM (s. Abb. 4a und b) zusammengefasst. Diese Aufteilung erfolgte in Anlehnung an Hyde et al. (Hyde et al. 1990). Eine alternative Aufteilung der PISA-Aufgaben nach mathematischen Stoffgebieten (Arithmetik, Algebra, Geometrie und Stochastik) lieferte vergleichbare Ergebnisse bezüglich von Geschlechtsunterschieden in Mathematik (s. Brunner 2005).

Zur Erfassung von IQ beziehungsweise IQ' wurden Figuren- (FA) und Wortanalogien (WA) aus dem kognitiven Fähigkeitstest (Heller und Perleth 2000) eingesetzt, die zu zwei Skalenscores zusammengefasst wurden. Bei Aufgaben aus der Skala *Figurenanalogien* werden Figurenanordnungen folgender Form vorgegeben (wobei die richtige Alternative aus einer von fünf Alternativen herauszufinden ist): Die erste Figur (z.B. kleiner Kreis) verhält sich zu einer zweiten (z.B. großer Kreis) wie die dritte (z.B. kleines Quadrat) zu einer von fünf Alternativen (in diesem Fall: großes Quadrat). Aufgaben aus der Skala *Wortanalogien* erfordern eine Ergänzung von Wortanalogien der Form A : B = C : ? (z.B. Huhn : Kücken = Kuh : ?). Aus fünf Alternativen ist wieder die korrekte Antwort (im Bsp: Kalb) auszuwählen. Diese beiden Skalen haben sich in vielen large-scale Untersuchungen als reliable und valide Maße bewährt.

Um mehr über die Relevanz von M und M' zu erfahren, synthetisieren wir in diesem Beitrag die Ergebnisse früherer Studien (Brunner 2005, 2008; Brunner und Krauss 2010; Brunner et al. 2008b), die ebenfalls die PISA 2000 Datenbasis nutzten. Diese Studien untersuchten den Zusammenhang von Geschlecht, M und M' einerseits und wichtigen Schülermerkmalen andererseits. In diesen Studien wurden beispielsweise die folgenden motivationalen Schülermerkmale analysiert (für eine genaue Beschreibung der eingesetzten Instrumente, s. Kunter et al. 2002): *Interesse an Mathematik* (Beispielitem: „Mathematik ist mir persönlich wichtig“), *Interesse am Lesen* (Beispielitem: „Weil mir das Lesen Spaß macht, würde ich es nicht gerne aufgeben“), *Selbstkonzept in Mathematik* (Beispielitem: „Ich war schon immer gut in Mathematik“), *Selbstkonzept in Deutsch* (Beispielitem: „Im Fach Deutsch lerne ich schnell“). Die Schülerinnen und Schüler konnten bei allen Items auf 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. Darüber hinaus wurden in diesen Studien auch die Korrelationen mit *Deutsch- und Mathematiknoten* betrachtet. Die Erfassung der Schulnoten basierte auf Schülerselbstberichten. Hierzu beantworteten die Schüler die Frage „Welche Zensuren hattest du im letzten Zeugnis?“ für die Fächer Mathematik und Deutsch mithilfe einer Notenskala (1 = „sehr gut“, 2 = „gut“, 3 = „befriedigend“, 4 = „ausreichend“, 5 = „mangelhaft“, 6 = „ungenügend“). Aufgrund der Orientierung der Notenskala bedeuten negative Korrelationen zwischen kognitiven Kompetenzen und Noten, dass höhere Kompetenzwerte mit besseren Noten einhergingen.

Als statistische Analysemethode zur Analyse der Geschlechtsunterschiede in Standardmodell und Nested-Faktormodell verwendeten wir konfirmatorische Mehrgruppen-Faktorenmodelle (z.B. Millsap und Kwok 2004). Als statistische Analysemethode zur Zusammenhangsanalyse mit Schülermerkmalen verwendeten wir konfirmatorische Faktorenanalysen (hierbei wurden motivationale Merkmale als latente Variablen und Schulnoten als manifeste Variablen repräsentiert). Detaillierte Informationen zu den verwendeten Instrumenten und zum statistischen Vorgehen finden sich in den Arbeiten von Brunner (2005) beziehungsweise Brunner et al. (2008a).

6 Ergebnisse

6.1 Geschlechtsunterschiede im Modellvergleich

Bevor die zentralen Ergebnisse referiert werden, muss überprüft werden, ob die beiden verwendeten Modelle die beobachteten Daten überhaupt adäquat beschreiben können. Sowohl das Standardmodell (Abb. 4a) als auch das Nested-Faktormodell (Abb. 4b) wiesen einen guten Modell-Fit auf (vergleiche hierzu Browne und Cudeck 1993; Hu und Bentler 1998). Wie die deskriptiven Fit-Indizes (CFI, RMSEA und SRMR) zeigen, konnte das Nested-Faktormodell die empirischen Relationen zwischen den Skalenscores etwas besser erklären als das Standardmodell.

Alle Faktorladungen waren substantiell und statistisch signifikant von Null verschieden. In Abb. 4a und b befinden sich jeweils an den Pfeilen oben die Faktorladung der Mädchen und unten diejenigen der Jungen. Die Faktorladungen geben dabei (analog zu Regressionsgewichten) an, wie stark der Zusammenhang zwischen den latenten Variablen und den manifesten Skalenscores ist: Je höher eine Faktorladung ist, desto stärker ist dieser Zusammenhang. Abbildung 4b legt daher auch nahe, dass es keine „reinen“ Mathematikskalen gibt. Auch wenn eine reine mathematische *Kompetenz* theoretisch vorstellbar und somit auch modellierbar ist, gibt es – zumindest bei PISA – keine Skala, die diese Kompetenz M' „in Reinform“ repräsentiert, denn jede Mathematikskala wurde über M' hinaus auch substantiell von IQ' beeinflusst. Weiterführende Analysen mittels der Faktorladungen zeigten, dass M' und IQ' in etwa ein beziehungsweise vier Fünftel der reliablen Gesamtvarianz der Mathematikskalen erklären konnten; das „Mischungsverhältnis“ der reliablen Varianzen lag also in etwa bei eins zu vier. Zusammen genommen können also *beide* Modelle die beobachteten Daten adäquat repräsentieren. Somit ist eine weitere Analyse der Geschlechtsunterschiede in beiden Modellen grundsätzlich sinnvoll.

Wie unterschieden sich die Mittelwerte in IQ, IQ', M und M' von Jungen und Mädchen? Erwartungsgemäß waren Geschlechtsunterschiede in der Intelligenz um

Null (Abb. 5). Mädchen erzielten geringfügig bessere Leistungen als Jungen in IQ ($d = -.08$) wie auch in IQ' ($d = -.09$). Weiterführende Analysen zeigten, dass dieser Intelligenzvorsprung der Mädchen damit zusammenhängt, dass sie häufiger als Jungen höhere Schulformen besuchen (Brunner et al. 2008a). Die annähernd gleichen Leistungsdifferenzen für IQ und IQ' unterstreichen auch noch einmal die postulierte Äquivalenz der beiden Konstrukte (siehe Abschn. 3).

Der wichtigste Befund war, dass unsere Schlussfolgerungen zu Geschlechtsunterschieden in M und M' bestätigt wurden. Die üblicherweise kleinen Geschlechtsunterschiede in M wurden bei Verwendung des Standardmodells repliziert; Der Leistungsvorsprung der Jungen in M betrug $d = .35$.

Hingegen zeigen die Ergebnisse für das Nested-Faktormodell, dass Jungen in der *mathematikspezifischen Kompetenz* M' deutlich besser waren als Mädchen ($d = .94$) (Abb. 5). Betrachtet man also eine „reine mathematische Kompetenz“, die auf Konstruktebene nicht mehr mit Intelligenz „vermischt“ ist, sind Geschlechtsunterschiede *wesentlich* größer als bei der üblichen Modellierung. Oder anders ausgedrückt: In Intelligenz gibt es keine Geschlechtsunterschiede. Die in Mathematik üblicherweise gefundenen Geschlechtsunterschiede sind deshalb so „gering“, weil zur Lösung von Mathematikaufgaben auch Intelligenz erforderlich ist.

Welche weiteren Erkenntnisse lassen sich – über die unterschiedliche Größe von Geschlechtsunterschieden hinaus – aus unseren Analysen ableiten?

6.2 Was kann man sich unter Geschlechtsunterschieden in M' vorstellen?

Die im vorherigen Abschnitt dargestellten Ergebnisse machen deutlich, dass das gewählte Messmodell mathematischer Kompetenz einen starken Einfluss auf die Größe der vorgefundenen Geschlechtsunterschiede in Mathematik hat. In diesem Abschnitt wollen wir der Frage nach der *Bedeutung von* M' nachgehen. Dazu betrachten wir zwei statistische Parameter: (1) Die Faktorladungen von Mathematikaufgaben, bzw. -skalenscores auf M', da diese darüber informieren, welche Maße am stärksten von M' beeinflusst werden. (2) Gleichzeitig nehmen wir diejenigen Mathematikaufgaben, bzw. -skalenscores unter die Lupe, die große Geschlechtsunterschiede zu Gunsten der Jungen aufweisen. Der Grund hierfür ist, dass in M' große Geschlechtsunterschiede gefunden wurden und diese sich daher in den Mathematikaufgaben, bzw. -skalenscores manifestieren sollten.

Betrachten wir dazu zuerst die Faktorladungen der Aufgabenklassen und der mathematischen Stoffgebiete auf M' sowie die Geschlechtsunterschiede in den entsprechenden Skalenscores (Tab. 2). Die Faktorladungen wurden dabei mit konfirmatorischen Faktorenanalysen auf Grundlage der Gesamtstichprobe berechnet.

Erwartungsgemäß zeigen sich Geschlechtsunterschiede zu Gunsten der Jungen bei den Modellierungsaufgaben und dort wiederum im Besonderen bei den rechnerischen Modellierungsaufgaben (vgl. Hyde et al. 1990; Walther et al. 2008). Gleichzeitig ist aus Tab. 2 ersichtlich, dass rechnerische und begriffliche Modellierungsaufgaben die höchsten Faktorladungen auf M' aufweisen.

In Bezug auf die Stoffgebiete finden sich (ebenfalls erwartungsgemäß) die größten Geschlechtsunterschiede in Geometrie (Tab. 2). Entgegen der Erwartung (siehe Abschn. 2) zeigen sich jedoch auch in Arithmetik vergleichbare Geschlechtsunterschiede, Arithmetik weist darüber hinaus sogar die höchste Faktorladung auf M' auf.

Tab. 2 Faktorladungen im Nested-Faktormodell und (modellunabhängige) Geschlechtsunterschiede in den manifesten Skalenscores für Aufgabenklassen und mathematische Stoffgebiete

Skalenscore	Faktorladung im Nested-Faktormodell auf M'	Geschlechts-unterschiede d
Aufgabenklassen		
Technische Aufgaben	.25	−.02
Rechnerisches Modellieren	.43	.31
Begriffliches Modellieren	.35	.22
Stoffgebiete		
Algebra	.31	.10
Arithmetik	.42	.24
Geometrie	.30	.26
Daten & Zufall	.25	.12

Anmerkung. Positive Werte der Effektgröße d bedeuten, dass Jungen bessere Leistungen in Mathematik erzielten als Mädchen

Es liegt deshalb nahe, auf der Suche nach der Natur von M' einen genaueren Blick auf die PISA-Aufgaben zu Arithmetik zu werfen. Mit Hilfe des im Rahmen der COACTIV-Studie entwickelten Aufgabenklassifikationsschemas (Jordan et al. 2006, 2008) wurden die PISA-Aufgaben noch einmal feiner unterteilt. Die Arithmetik-Aufgaben wurden hierzu in die Unterbereiche *Proportionalität und Antiproportionalität*, *Potenzen und Wurzeln*, *Prozent- und Zinsrechnung* sowie *Zahlbereiche und Rechnen* klassifiziert. Abbildung 6 zeigt auf der y-Achse die Geschlechtsunterschiede (positive Werte dieser Effektgröße indizieren dabei Vorteile für die Jungen). Jeder Punkt stellt dabei eine einzelne Arithmetikaufgabe dar. Auf der x-Achse sind die (standardisierten) Faktorladungen der Arithmetikaufgaben auf M' abgetragen. Diese Faktorladungen wurden hierbei mittels eines nonlinearen Item-Faktormodells ermittelt (Muthén und Muthén 1998–2010). In diesem Modell wurden alle Mathematikaufgaben durch IQ' und M' beeinflusst. Intelligenztestaufgaben wurden durch IQ' und jeweils durch eine korrespondierende (latente) aufgabenspezifische Kompetenz (für Figuren-, bzw. Wortanalogien) beeinflusst (dies entspricht dem Rest-Term in Abb. 2d). Alle latenten Kompetenzen wurden als wechselseitig unkorreliert spezifiziert. M' und IQ' in diesem Modell können somit analog zum Nested-Faktormodell, das in Abb. 4a dargestellt ist, interpretiert werden.

Aus Abb. 6 ist ersichtlich, dass (in Bezug auf Arithmetik) M' vornehmlich von Aufgaben zur Prozent- und Zinsrechnung repräsentiert wird: Diese Aufgaben laden am höchsten auf M' und zeigen gleichzeitig auch die größten Geschlechtsunterschiede zu Gunsten der Jungen.

Ähnliche differentielle Befunde zeigen sich für Algebra. Bei Aufgaben zu den Untergebieten quadratische Gleichungen und quadratische Funktionen ergeben sich relativ hohe Faktorladungen auf M' (im Mittel jeweils .34), bei Aufgaben zu Termen, Variablen und linearen Gleichungen/Funktionen etwas niedrigere (die mittlere Faktorladung lag hier bei .24). Die Geschlechtsunterschiede in den Algebraaufgaben – auch bei denen zu quadratischen Gleichungen und Funktionen – sind dabei aber eher moderat (quadratische Gleichungen und Funktionen: mittlere Effektgröße = .09; Terme, Variablen und lineare Gleichungen/Funktionen: mittlere Effektgröße = .06). Für die Geometrie und für die Stochastik ergeben sich keine vergleichbar auffälligen differentiellen Befunde. Während Aufgaben zu quadratischen Gleichungen und quadratischen Funktionen also stark von M' beeinflusst werden (aber nur geringfügig höhere Geschlechtsunterschiede als die anderen Aufgaben aus der Algebra aufweisen), trifft für Aufgaben aus der Prozentrechnung beides zu: Sie laden hoch auf M' *und* zeigen sehr starke Geschlechtsunterschiede. Auf die Bedeutung der Prozentrechnung in der Forschung zu Geschlechtsunterschiedenen in Mathematik gehen wir in Abschn. 7 noch einmal ein.

6.3 Der Zusammenhang von M und M' mit Schülermerkmalen

Der vorherige Abschnitt gab erste Hinweise zur inhaltlichen Bedeutung von Geschlechtsunterschieden in M'. Um mehr über die generelle Relevanz von M' zu erfahren, fassen wir in diesem Abschnitt Ergebnisse früherer Studien (Brunner 2005, 2008; Brunner und Krauss 2010; Brunner et al. 2008b) zum Zusammenhang von M und M' einerseits und wichtigen Schülermerkmalen andererseits zusammen, die wir mit Ergebnissen zu Geschlechtsunterschieden in diesen Schülermerkmalen ergänzen.

Abbildung 7 stellt einige zentrale Ergebnisse dar. In dieser Abbildung sind die Korrelationen zwischen Schülermerkmalen und M' als graue Balken dargestellt; die entsprechenden Korrelationen für M sind als weiße Balken repräsentiert. Ebenso sind in Abb. 7 die Mittelwertsunterschiede zwischen und Jungen und Mädchen in diesen Schülermerkmalen als schraffierte Balken dargestellt. Diese Mittelwertsunterschiede sind hierbei als punktbiserale Korrelationen mit der Variablen Geschlecht (Mädchen wurden mit 0, Jungen mit 1 kodiert) repräsentiert; diese Korrelationen haben wir zur besseren Verständlichkeit auch als d-Werte ausgedrückt (s. Cohen 1988, S. 23).

Betrachten wir zunächst die Geschlechtsunterschiede. Jungen hatten (im Vergleich zu Mädchen) tendenziell (1) ein höheres mathematisches Selbstkonzept von sich, (2) waren interessierter an Mathematik und (3) erzielten bessere Noten in Mathematik. Andererseits hatten Jungen im Vergleich zu Mädchen, (4) tendenziell ein niedrigeres verbales Selbstkonzept, (5) waren weniger interessiert am Lesen und (6) erzielten schlechtere Deutschnoten.

Betrachten wir nun die Zusammenhänge zwischen M und M' und den Schülermerkmalen. Abbildung 7 zeigt ein Befundmuster auf, das Gemeinsamkeiten, aber auch sehr deutliche Unterschiede zwischen M und M' identifiziert. Die Gemeinsamkeit von M und M' betraf, dass Schülerinnen und Schüler mit höheren Werten auf M beziehungsweise auf M' tendenziell (1) ein höheres mathematisches Selbstkonzept von sich besaßen, (2) interessierter an Mathematik waren und (3) bessere Noten in Mathematik erzielten.

Der zentrale Unterschied zwischen M und M' betrifft ihre Zusammenhänge mit sprachlichen Schülermerkmalen. Aus Abb. 7 ist ersichtlich, dass Schülerinnen und Schüler mit höheren Werten auf M tendenziell auch (4) ein höheres verbales Selbstkonzept von sich hatten, (5) interessierter am Lesen waren und (6) bessere Deutschnoten erzielten. Für M' resultierte jedoch ein deutlich anderes Befundmuster. Schülerinnen und Schüler mit höheren Werten auf M' (4) hatten tendenziell ein *geringeres* verbales Selbstkonzept von sich, (5) waren *weniger* interessiert am Lesen und (6) erzielten *schlechtere* Deutschnoten.

Insgesamt verdeutlichen die Ergebnisse für das Nested-Faktormodell, dass die mathematikspezifische Kompetenz M' positiv mit mathematischen Schülermerkmalen und – im Gegensatz zu M – negativ mit sprachlichen Schülermerkmalen zusammenhängt. Wie Abb. 7 klar zeigt, ist dieses Befundmuster kongruent zu den Geschlechtsunterschieden in den untersuchten Schülermerkmalen. Hierbei ist es wichtig hervorzuheben, dass dieses Befundmuster wie auch die gefundene Kongruenz keineswegs trivial ist. *Nur unter Verwendung des Nested-Faktormodells* zeigen sich diese theoretisch auch zu erwartenden (siehe dazu Abschn. 7) differentiellen Zusammenhänge (4 bis 6 in Abb. 7) der mathematikspezifischen Kompetenz M'. Bei der Verwendung des Standardmodells konnten diese differentiellen Zusammenhänge sowie die Kongruenz zu den Geschlechtsunterschieden hinsichtlich von Schülermerkmalen im mathematischen und sprachlichen Bereich *nicht* gefunden werden. Die Implikationen dieser Ergebnisse für das Verständnis von schulischen Entwicklungsprozessen diskutieren wir im folgenden Abschnitt.

7 Diskussion

Die wichtigsten Ergebnisse unseres Artikels können wie folgt zusammengefasst werden: Erstens, die Größe von Geschlechtsunterschieden in Mathematik hängt stark vom verwendeten Messmodell „mathematischer Kompetenz“ ab. Zweitens, Aufgaben, die Prozentrechnen erfordern, sind die Mathematikaufgaben, die gleichzeitig am stärksten von M' beeinflusst werden und große Geschlechtsunterschiede zu Gunsten der Jungen aufweisen. Drittens, M und M' zeigen ein ähnliches Zusammenhangsmuster mit Lernmotivation und Noten in Mathematik, unterscheiden sich aber deutlich darin, wie sie mit der Lernmotivation und Noten im Fach Deutsch korrelieren. Wir diskutieren diese Befunde nun vor dem Hintergrund von drei Fragen.

(1) *Ist eine Neubewertung von Geschlechtsunterschieden in Mathematik notwendig?* In diesem Artikel haben wir Geschlechtsunterschiede in Mathematik mit Hilfe von zwei alternativen, theoretisch gestützten Messmodellen mathematischer Kompetenz – dem Standardmodell und dem Nested-Faktormodell – auf Grundlage der Daten der deutschen PISA-2000 Stichprobe analysiert. Im in der Bildungsforschung (aber auch generell in der pädagogischen und psychologischen Forschung) vorherrschenden „Standardmodell“ zeigten sich in unseren Analysen erwartungsgemäß eher geringe ($d = .35$) Geschlechtsunterschiede zu Gunsten der Jungen. Hingegen fanden sich mit dem in der Intelligenzforschung präferierten Nested-Faktormodell beim selben Datensatz sehr große ($d = .94$) Geschlechtsunterschiede zu Gunsten der Jungen in einer mathematikspezifischen Kompetenz M' (dieser Befund steht damit im Einklang mit den Ergebnissen von Rosén 1995). Da sowohl das Standardmodell als auch das Nested-Faktormodell theoretisch und empirisch gestützt werden, hängt die Frage nach der Größe von Geschlechtsunterschieden also von einer theoretischen Annahme ab: Beeinflusst die generelle kognitive Fähigkeit (die Intelligenz) die Leistung bei Mathematikaufgaben (Nested-Faktormodell) oder nicht (Standardmodell)? Beantwortet man diese Frage mit „ja“, sind Geschlechtsunterschiede in Mathematik wesentlich größer als bisher angenommen.

(2) *Was ist M'*? Die mathematikspezifische Kompetenz M' ist ein latentes (also nicht direkt beobachtbares) Konstrukt, das man sich als einen Blick durch eine Lupe, in der alle Intelligenzanteile herausgefiltert werden, auf das (ebenfalls latente) Konstrukt M vorstellen kann. Die Suche nach manifesten (also beobachtbaren Indikatoren) für M' gestaltet sich jedoch als schwierig, da die PISA-Mathematikaufgaben im Sinne des Standardmodells konzipiert wurden. Vor allem Aufgaben zur Prozentrechnung und zu quadratischen Gleichungen und quadratischen Funktionen scheinen aber – zumindest was die PISA-Aufgaben betrifft – dem Kern von M' nahe zu stehen. Es ist aber durchaus vorstellbar, das neue Aufgaben gefunden (bzw. konzipiert) werden können, die noch höhere Faktorladungen auf M' (und gleichzeitig noch niedrigere auf IQ') aufweisen. Es ist eine Aufgabe für zukünftige Forschung solche Aufgabentypen zu identifizieren, denn es könnte sein, dass diese Aufgaben neue Erkenntnisse über Geschlechtsunterschiede liefern.

So wurden wir durch unsere Analysen auf einen Bereich aufmerksam, in dem sich – neben hohen Faktorladungen auf M' – gleichzeitig auch relativ große Geschlechtsunterschiede zeigen, nämlich die Prozentrechnung. Es ist bekannt, dass inverse Prozentaufgaben für Schülerinnen *und* Schüler generell schwieriger sind als direkte. Weiterhin deuten einige empirische Studien daraufhin, dass es insbesondere bei inversen Prozentrechenaufgaben relativ große Geschlechtsunterschiede zu Gunsten der Jungen gibt (Martignon 2009).

Warum wurden Aufgaben zur Prozentrechnung bislang in größeren Studien nicht als geschlechtssensitive Aufgaben identifiziert? Mindestens zwei Erklärungsmöglichkeiten bieten sich an. Erstens könnte einer der Gründe darin liegen, dass bestimmte Aufgaben zum Prozentbegriff bei Aufgabenklassifizierungen – je nach Anleitung im Kodiermanual – einmal zu Daten und Zufall (wie bei PISA) und einmal zu Arithmetik (wie z.B. im COACTIV-Schema zur Aufgabenklassifikation, in dem Prozentrechnen als eigenständige Subkategorie der Arithmetik aufgeführt wird) gerechnet werden können. In der Tat waren fünf der Prozentaufgaben aus Abb. 6 bei PISA der Stochastik zugeordnet. Da die Ergebnisse in PISA (aber auch in den Meta-Analysen von Janet Hyde und Kolleginnen oder in TIMSS) nicht so detailliert aufgeschlüsselt wurden wie in der vorliegenden Arbeit, wurden die relativ großen Geschlechtsunterschiede im Prozentrechnen vermischt mit relativ kleinen Geschlechtsunterschieden in anderen Teilbereichen der beiden jeweiligen mathematischen Stoffgebiete. Zweitens, in PISA (oder TIMSS) werden durch großangelegte empirische Pilotierungsstudien vorab die Aufgaben identifiziert, die über generelle Geschlechtseffekte in der mathematischen Kompetenz M hinaus noch größere Geschlechtsunterschiede zeigen. Insbesondere führten die Analysen von differentiellen Itemfunktionen (sogenannte DIF-Analysen) dazu, dass die Aufgaben, die am stärksten geschlechtssensitiv waren, letztlich nicht in den Mathematiktest für die Hauptstudie von PISA 2000 aufgenommen wurden (Adams und Wu 2002). Dies lässt vermuten, dass die gefundenen Effekte bei den Prozentrechnungsaufgaben in PISA 2000 eher die Untergrenze für Geschlechtsunterschiede beim Prozentrechnen darstellen als die Obergrenze. Für Bemühungen zum Ausgleich von Geschlechtsunterschieden (Budde 2009; Heinze et al. 2007; Leder und Forgasz 2008; Martignon et al. 2006; Zentralblatt für Didaktik der Mathematik 1994) sollten Aufgaben, die Prozentrechnungen benötigen, jedenfalls zukünftig verstärkt mit in den Blick genommen werden (z.B. durch gezielte qualitative Analysen des lauten Denkens, die Aufschluss geben können, weshalb diese Aufgaben Mädchen besonders schwer fallen).

(3) *Wie beeinflusst M' die Entwicklung von Schülerinnen und Schülern*? Die Investition in den Erwerb mathematischen Wissens sollte mit gesteigerter Lernmotivation und auch mit besseren Noten in Mathematik einhergehen (vgl. Ackerman 1996). Aufgrund begrenzter zeitlicher sowie kognitiver und motivationaler Ressourcen von Schülerinnen und Schülern sollte die kognitive und motivationale Entwicklung in einem Unterrichtsfach potenziell zu Lasten der kognitiven und motivationalen Entwicklung in anderen Unterrichtsfächern gehen. Diese theoretisch plausible (und qualitativ oft berichtete) Annahme der Fokussierung auf *entweder* Mathematik *oder* Deutsch kann erstaunlicherweise mit M im Rahmen des Standardmodells empirisch nicht nachgewiesen werden, wohl aber mit M' (Abb. 7). Das Befundmuster von M' im Nested-Faktormodell ist dabei kongruent mit dem Befundmuster zu Geschlechtsunterschieden in mathematischen und sprachlichen Schülermerkmalen (s. Abb. 7). Insgesamt zeigen diese Befunde zur Konstruktvalidität mit externen Variablen, dass sich verstärkte Forschung zu M' mittels des Nested-Faktormodells lohnt. Weiterführende Analysen zur diskriminanten und konvergenten Validität von M und M' in Bezug auf Schulformunterschiede und sozio-ökonomischen Hintergrund finden sich in Brunner (2005).

Gegenwärtig sind Männer in Berufen in mathematisch-naturwissenschaftlich-technischen Bereichen deutlich überrepräsentiert (Ceci et al. 2009; Organisation for Economic Co-operation and Development 2009). Bislang wurden hierfür viele Erklärungsansätze diskutiert, die das komplexe Zusammenwirken von biologischen, sozialen, pädagogischen, motivationalen und kognitiven Faktoren einbeziehen (z.B. Budde 2009; Ceci et al. 2009; Else-Quest et al. 2010; Leder und Forgasz 2008). Wichtig ist hierbei aber zu betonen, dass die große Relevanz der psychometrischen Modellierung mathematischer Kompetenz, insbesondere eine Modellierung der mathematikspezifischen Kompetenz mit dem Nested-Faktormodell, für die Bewertung von Geschlechtsunterschieden in diesen Erklärungsansätzen bislang keine Beachtung fand. Wir möchten daher diese Erklärungsmodelle um eine psychometrisch-basierte Hypothese erweitern: Da Kompetenzen und die fachspezifische Lernmotivation zu den bedeutsamsten Prädiktoren der späteren Berufswahl zählen, spekulieren wir, dass vor allem Schülerinnen und Schüler mit höheren Werten auf M' später Berufe im mathematisch-naturwissenschaftlich-technischen Bereich und nicht etwa im sprachlichen Bereich wählen werden. Da in M' große Geschlechtsunterschiede zu Gunsten der Jungen gefunden wurden, könnte M' *eine* zentrale Erklärungsvariable sein, warum Männer in mathematisch-naturwissenschaftlich-technischen Berufen deutlich überrepräsentiert sind. Wir gehen dabei davon aus, dass M' seine Wirkung in reziproker Interaktion mit motivationalen Variablen (z.B. Selbstkonzept und Interesse) entfaltet. Eine Überprüfung dieser Hypothesen zur schulischen und beruflichen Entwicklung auf Grundlage repräsentativer Längsschnittstichproben steht noch aus, könnte aber beispielsweise mit den Daten aus der deutschen National Educational Panel Study (NEPS) realisiert werden.

Schließlich ist es uns wichtig festzuhalten, dass wir *nicht* über mögliche biologisch-genetische Erklärungen spekulieren, wie die Geschlechterunterschiede in M' zustande kommen. Dies wollen wir insbesondere deshalb auch nicht tun, da einerseits die zum Teil deutlichen Unterschiede in Geschlechtsunterschieden in Mathematik zwischen Ländern gegen eine universale biologisch-genetische Erklärung sprechen (Else-Quest et al. 2010) und andererseits, weil die Arbeiten von Guiso et al. (2008) und von Else-Quest et al. (2010) auf die Korrelation zwischen Gleichstellungsfaktoren und mathematischer Leistung von Mädchen in Mathematik aufmerksam machten. Zukünftige Forschung sollte demnach vor allem der Frage nachgehen, welche sozialen und kulturellen Faktoren die mathematikspezifische Kompetenz M' fördern.

Literatur

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence*, *22*, 227–257.

Adams, R., & Wu, M. (2002). *PISA 2000 technical report*. Paris: OECD.

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.

Beaton, A. E., Martin, M. O., Mullis, I. V. S., Gonzales, E. J., Smith, T. A., & Kelly, D. L. (1996). *Mathematics achievement in the middle school years*. Chestnut Hill: TIMSS International Study Center, Boston College.

Beller, M., & Gafni, N. (Hrsg.) (1996). The 1991 International assessment of educational progress in mathematics and sciences: The gender differences perspective. *Journal of Educational Psychology*, *88*, 365–377.

Blum, W., Drüke-Noe, C., Hartung, R., & Köller, O. (Hrsg.) (2006). *Bildungsstandards Mathematik: konkret Sekundarstufe I. Aufgabenbeispiele, Unterrichtsanregungen, Fortbildungsideen*. Berlin: Cornelsen-Scriptor-Verlag.

Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *110*, 305–314.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Hrsg.), *Testing structural equation models* (S. 136–162). Newbury Park: Sage.

Brunner, M. (2005). *Mathematische Schülerleistung: Struktur, Schulformunterschiede und Validität*. Dissertation, Humboldt-Universität zu Berlin. http://edoc.huberlin.de/dissertationen/brunner-martin-2006-02-08/PDF/brunner.pdf. Gesehen 7. Juni 2006.

Brunner, M. (2008). No *g* in education? *Learning and Individual Differences*, *18*, 152–165.

Brunner, M., & Krauss, S. (2010). Modellierung kognitiver Kompetenzen von Schülern und Lehrkräften mit dem Nested-Faktormodell. In W. Bos, E. Klieme, & O. Köller (Hrsg.), *Schulische Lerngelegenheiten und Kompetenzentwicklung. Festschrift für Jürgen Baumert* (S. 105–126). Münster: Waxmann.

Brunner, M., Krauss, S., & Kunter, M. (2008a). Gender differences in mathematics: Does the story need to be rewritten? *Intelligence*, *36*, 403–421.

Brunner, M., Lüdtke, O., & Trautwein, U. (2008b). The internal/external frame of reference model revisited: Incorporating general cognitive ability and general academic self-concept. *Multivariate Behavioral Research*, *43*, 137–172.

Budde, J. (2009). *Mathematikunterricht und Geschlecht. Empirische Ergebnisse und pädagogische Ansätze*. Berlin: Bundesministerium für Bildung und Forschung (BMBF).

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.

Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin*, *135*, 218–261.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Lawrence Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3. Aufl.). Mahwah: Erlbaum.

Eid, M., Lischetzke, T., Nussbeck, F. W., & Trierweiler, L. I. (2003). Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple-indicator CT-C(M-1) model. *Psychological Methods*, *8*, 38–60.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, *136*, 103–127.

Exekutivagentur Bildung (EACEA P9 Eurydice) (2010). *Geschlechterunterschiede bei Bildungsresultaten: Derzeitige Situation und aktuelle Maßnahmen in Europa*. Brüssel: EACEA P9 Eurydice.

Gallagher, A. M., De Lisi, R., Holst, P. C., McGillicuddy-De Lisi, A. V., Morely, M., & Cahalan, C. (2000). Gender differences in advanced mathematical problem solving. *Journal of Experimental Child Psychology*, *75*, 165–190.

Gallagher, A. M., & Kaufman, J. C. (Hrsg.) (2005). *Gender differences in mathematics*. New York: Cambridge University Press.

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history and bibliography. *Intelligence*, *24*, 13–23.

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *Science*, *320*, 1164–1165.

Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, *8*, 179–203.

Gustafsson, J. E. (1988). In R. J. Sternberg (Hrsg.), *Advances in the psychology of human intelligence: Bd. 4. Hierarchical models of individual differences in cognitive abilities* (S. 35–72). Hillsdale: Erlbaum.

Gustafsson, J. E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, *28*, 407–434.

Gustafsson, J. E., & Undheim, J. O. (1996). Individual differences in cognitive functions. In D. C. Berliner & R. C. Calfee (Hrsg.), *Handbook of educational psychology* (S. 186–242). New York: Macmillan.

Halpern, D. F., & LaMay, M. L. (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review*, *12*, 229–246.

Hartig, J., & Höhler, J. (2008). Representation of competencies in multidimensional IRT models with within-item and between-item multidimensionality. *Journal of Psychology*, *216*, 89–101.

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, *269*, 41–45.

Heinze, A., Kessler, S., Kuntze, S., Lindmeier, A., Moormann, M., Reiss, K. et al. (2007). Kann Paul besser argumentieren als Marie? Betrachtungen zur Beweiskompetenz von Mädchen und Jungen aus differentieller Perspektive. *Journal für Mathematik-Didaktik*, *28*(2), 148–167.

Heller, K. A., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision. Manual*. Göttingen: Hogrefe.

Hu, L.-T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424–453.

Hunt, E., & Carlson, J. (2007). Considerations relating to the study of group differences in intelligence. *Perspectives on Psychological Science*, *2*, 194–213.

Hyde, J. S. (2005). The gender similarity hypothesis. *The American Psychologist*, *60*, 581–592.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: a meta-analysis. *Psychological Bulletin*, *107*, 139–155.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science*, *321*, 494–495.

Jarvis, C. B., MacKenzie,S. B., & Podsakoff, P. M.(2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, *30*, 199–218.

Jensen, A. R. (1998). *The g factor. The science of mental ability*. Westport: Praeger.

Jordan, A., Krauss, S., Löwen, K., Blum, W., Neubrand, M., Brunner, M., et al. (2008). Aufgaben im COACTIV-Projekt: Zeugnisse des kognitiven Aktivierungspotentials im deutschen Mathematikunterricht. *Journal für Mathematik-Didaktik*, *29*(2), 83–107.

Jordan, A., Ross, N., Krauss, S., Baumert, J., Blum, W., Neubrand, M., et al. (2006). *Klassifikationsschema für Mathematikaufgaben: Dokumentation der Aufgabenkategorisierung im COACTIV-Projekt (Materialien aus der Bildungsforschung Nr. 81)*. Berlin: Max-Planck-Institut für Bildungsforschung.

Kaiser, G., & Steisel, T. (2000). Results of an analysis of the TIMS study from a gender perspective. *Zentralblatt für Didaktik der Mathematik*, *32*(1), 18–24.

Kunter, M., Schümer, G., Artelt, C., Baumert, J., Klieme, E., Neubrand, M. et al. (2002). *PISA 2000: Dokumentation der Erhebungsinstrumente*. Berlin: Max-Planck-Institut für Bildungsforschung.

Leder, G., & Forgasz, H. (2008). Mathematics education: new perspectives on gender. *ZDM – The International Journal on Mathematics Education*, *40*(4), 601–616.

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin*, *136*, 1123–1135.
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.
Lynn, R. (1999). Sex differences in intelligence and brain size: A developmental theory. *Intelligence*, *27*, 1–12.
Lynn, R., & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence*, *32*, 481–498.
Martignon, L. (2009). Mädchen und Mathematik. In I. Wyrobnik & A. Matzer (Hrsg.), *Handbuch für Mädchenpädagogik* (S. 148–158). Weinheim: Beltz Verlag.
Martignon, L., Niederdrenk-Felgner, C., & Vogel, R. (Hrsg.) (2006). *Mathematik und Gender. Berichte der Arbeitstagungen 2003–2005 und Beiträge des Arbeitskreises Frauen und Mathematik in der Gesellschaft für Didaktik der Mathematik e.V.* Hildesheim: Franzbecker.
Millsap, R. E., & Kwok, O. M. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods*, *9*, 93–115.
Mullis, I. V. S., Martin, M. O., & Foy, P. (2008). *TIMSS 2007. International mathematics report. Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. Chestnut Hill: TIMSS & PIRLS International Study Center.
Mullis, I. V. S., Martin, M. O., Gonzales, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003. International mathematics report. Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. Chestnut Hill: TIMSS & PIRLS International Study Center.
Mullis, I. V. S., Martin, M. O., Gonzales, E. J., Gregory, K. D., Garden, R. A., O'Connor, K. M., et al. (2000). *TIMSS 1999. International mathematics report. Findings from IEA's repeat of the third international mathematics and science study at the eighth grade*. Chestnut Hill: International Study Center.
Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus user's guide* (6 Aufl.). Los Angeles: Muthén & Muthén.
Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *The American Psychologist*, *51*, 77–101.
Neubrand, M. (2003). „Mathematical literacy"/„Mathematische Grundbildung". Der Weg in die Leistungstests, die mathematikdidaktische Bedeutung, die Rolle als Interpretationshintergrund für den PISA-Test. *Zeitschrift für Erziehungswissenschaft*, *6*, 338–356.
Neubrand, M., Biehler, R., Blum, W., Cohors-Fresenborg, E., Flade, L., Knoche, N., et al. (2001). Grundlagen der Ergänzung des internationalen PISA-Mathematik-Tests in der deutschen Zusatzerhebung. *Zentralblatt für Didaktik der Mathematik*, *33*(1), 45–59.
Organisation for Economic Co-operation and Development (2001). *Knowledge and skills for life: first results from the OECD programme for international student assessment (PISA) 2000*. Paris: OECD.
Organisation for Economic Co-operation and Development (2004). *Learning for tomorrow's world: first results from PISA 2003*. Paris: OECD.
Organisation for Economic Co-operation and Development (2007). *Analysis: Bd.* 1114. *PISA 2006. Science competencies for tomorrow's world*. Paris: OECD.
Organisation for Economic Co-operation and Development (2009). *Equally prepared for life? How 15-year-old boys and girls perform in school*. Paris: OECD.
Organisation for Economic Co-operation and Development (2010). *PISA 2009 results: what students know and can do – student performance in reading, mathematics, and science* (Bd. 1). Paris: OECD.
Rosén, M. (1995). Gender differences in structure, means and variances of hierarchically ordered ability dimensions. *Learning and Instruction*, *5*, 37–62.
Schwank, I., Armbrust, S., & Libertus, M. (2003). Prädikative versus funktionale Denkvorgänge beim Konstruieren von Algorithmen. *Zentralblatt für Didaktik der Mathematik*, *35*(3), 79–85.
Snow, R. E., & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. L. Linn (Hrsg.), *Educational measurement* (3. Aufl., S. 263–331). New York: American Council on Education and Macmillan Publishing Company.
Spearman, C. (1904). "General intelligence", objectively determined and measured. *The American Journal of Psychology*, *15*, 201–293.
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.

van der Sluis, S., Dolan, C. V., & Stoel, R. D. (2005). A note on testing perfect correlations in SEM. *Structural Equation Modeling*, *12*, 551–577.

van der Sluis, S., Posthuma, D., Dolan, C. V., de Geus, E. J. C., Colom, R., & Boomsma, D. I. (2006). Sex differences on the Dutch WAIS-III. *Intelligence*, *34*, 273–289.

Walther, G., Schwippert, K., Lankes, E.-M., & Stubbe, T. C. (2008). Können Mädchen doch rechnen? Vertiefende Analysen zu Geschlechtsdifferenzen im Bereich Mathematik auf Basis der Internationalen Grundschul-Lese-Untersuchung IGLU. *Zeitschrift für Erziehungswissenschaft*, *11*, 30–46.

Winkelmann, H., van den Heuvel-Panhuizen, M., & Robitzsch, A. (2008). Gender differences in the mathematics achievements of German primary school students: results from a German large-scale study. *ZDM – The International Journal on Mathematics Education*, *40*, 601–616.

Yung, Y.-F., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika*, *64*, 113–128.

Zentralblatt für Didaktik der Mathematik (1994). Analysen zum Thema Frauen und Mathematik [Themenheft] (1994). *Zentralblatt für Didaktik der Mathematik, 26*(1–2).

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.