

# Psychologie in Erziehung und Unterricht

Zeitschrift für  
Forschung und  
Praxis

53. Jahrgang 2006

Herausgeber  
Gottfried Spangler  
Ernst Hany  
Olaf Köller  
Sabine Walper

Ernst Reinhardt Verlag München Basel

# Klassen- und Schulunterschiede im Ausmaß von Geschlechterdiskrepanzen bei Leistungsbewertungen und Leistungsaspirationen: Ergebnisse einer Mehrebenenanalyse

Markus Dresel, Heidrun Stöger, Albert Ziegler  
Universität Ulm

## *Classroom and School Differences Relating the Extent of Gender Disparities in Achievement Assessments and Achievement Aspirations: Results from a Multilevel Analysis*

**Summary:** With a sample of 9118 pupils attending 498 classes between the 5th and 10th grades at 54 schools we conducted gender differences in achievement assessments (report card grades) and achievement aspirations with a multilevel analysis, including individual, classroom and school level. Advantages were verified among boys in the subject of Mathematics, and among girls in the subject of German (native language) and in foreign language courses. For all three subjects, we identified substantial variations among classrooms regarding the extent of these gender differences. Furthermore, evidence of a widening gap in a temporal sense could be proved for German and foreign language instruction, whereby the advantages displayed by the girls expand with higher grade levels. Moreover, evidence of a widening gap in a vertical sense could be proved for German and foreign language courses, whereby these gender differences are less substantial in college preparatory schools than in other school types.

**Keywords:** Gender differences, scholastic achievement, level of aspiration, classroom differences, school differences

**Zusammenfassung:** An einer Stichprobe von 9118 Schüler(inne)n aus 498 Klassen der 5. bis 10. Jahrgangsstufe an 54 koedukativen Schulen wurden Geschlechtsunterschiede bei Leistungsbewertungen (Zensuren) und Leistungsaspirationen unter Einbezug von Individual-, Klassen- und Schulebene analysiert. Nachgewiesen wurden bessere Werte von Jungen im Mathematik- und bessere Werte von Mädchen im Deutsch- und Fremdsprachenunterricht. In allen drei Schulfächern konnten substanzielle Unterschiede zwischen den untersuchten Schulklassen im Ausmaß der Geschlechterdiskrepanzen identifiziert werden. Daneben wurde für den Deutsch- und den Fremdsprachenunterricht Evidenz erbracht sowohl für einen temporalen Schereneffekt, wonach die Geschlechtsunterschiede zugunsten von Mädchen mit zunehmender Jahrgangsstufe an Substanz gewinnen, als auch für einen vertikalen Schereneffekt, wonach diese Unterschiede in Haupt- und Realschule stärker ausgeprägt sind als am Gymnasium.

**Schlüsselbegriffe:** Geschlechtsunterschiede, Schulleistung, Aspirationsniveau, Klassenunterschiede, Schulunterschiede

Geschlechtsunterschiede im schulischen Leistungsbereich sind ein seit langer Zeit diskutiertes Problemfeld (für einen Überblick s. Hannover, 2004). Breit beforscht ist insbesondere der mathematisch-naturwissenschaftliche Bereich, wobei sich Unterschiede zu Ungunsten der Mädchen ab der Sekundarstufe zeigen, und zwar weitgehend unabhängig davon, ob Leistungsbewertungen durch Lehr-

kräfte (Zensuren) oder kriteriale Schulleistungstests Verwendung finden (z. B. Baumert, Bos & Lehmann, 2000; Middleton & Spanias, 1999; Stanat & Kunter, 2001; Tiedemann & Faber, 1994). Daneben finden sich in mathematisch-naturwissenschaftlichen Fächern auch Motivationsunterschiede: Im Mittel haben hier Mädchen ein geringeres Fachinteresse, geringere Erfolgserwartungen und ein niedri-

geres Fähigkeitsselfkonzept als Jungen (z. B. Dresel, Heller, Schober & Ziegler, 2001; Beyer & Bowden, 1997; Hannover, 1991; Hoffmann, Häußler & Peters-Haft, 1997; Rustemeyer & Jubel, 1996; Ziegler, Dresel & Schober, 2000). Darüber hinaus verfügen sie durchschnittlich über einen weniger motivations- und selbstwertförderlichen Attributionsstil – sie führen Erfolge weniger stark auf ihre eigenen Fähigkeiten sowie stärker auf personexterne Faktoren zurück und erklären sich Misserfolg stärker durch Fähigkeitsdefizite als Jungen (z. B. Dresel et al., 2001; Ziegler et al., 2000). Als Folge dieser Unterschiede haben Mädchen in mathematisch-naturwissenschaftlichen Domänen niedrigere Leistungsaspirationen, geben sich also mit schlechteren Leistungsergebnissen zufrieden als Jungen (zsf. Beerman, Heller & Menacher, 1992).

Die internationalen Schulleistungsstudien PISA 2000 und PISA 2003 bestätigten für den mathematischen Kompetenzbereich einen Vorteil der Jungen (Stanat & Kunter, 2001, 2003; Zimmer, Burba & Rost, 2004). Gleichzeitig rückten diese Studien auch mögliche Benachteiligungen von Jungen im sprachlichen Bereich in den Fokus der Diskussion. So erbrachte PISA 2003 Vorteile der Mädchen bei der Lesekompetenz (Zimmer et al., 2004), wobei die Effektgröße des Unterschieds im OECD-Durchschnitt bei  $d=0.35$  lag (Deutschland:  $d=0.39$ ). Im Vergleich dazu fiel der Vorteil der Jungen in der mathematischen Kompetenz mit  $d=0.11$  im OECD-Durchschnitt (Deutschland:  $d=0.09$ ) deutlich geringer aus. Daneben zeigte sich, dass das Ausmaß der Unterschiede in den 29 OECD-Staaten sehr stark variiert: Nach Zimmer et al. (2004) schwankt der Vorteil der Mädchen bei der Lesekompetenz zwischen  $d=0.61$  (Island) und  $d=0.21$  (Japan). Signifikante Vorteile der Jungen im mathematischen Kompetenzbereich werden nur für 21 Staaten berichtet; die Differenzen schwanken zwischen einem Vorteil der Jungen von  $d=0.26$  (Korea) und einem nahezu ebenso großen Vorteil der Mädchen von  $d=0.17$  (Island).

### *Sozialisationsinstanz Schulklasse*

Die meisten theoretischen Ansätze zur Erklärung der genannten Geschlechtsunterschiede heben auf eine geschlechtsspezifische Sozialisation ab (z. B. Halpern, 2000; Trautner, 1994; vgl. aber auch Eagly, Beall, & Sternberg, 2004). Viele Modelle enthalten die Annahme, dass geschlechtsspezifische Rollenerwartungen durch verschiedene Sozialisationsinstanzen vermittelt und kristallisiert als differenzielle Motivationen lern- und leistungsrelevant werden (z. B. Eccles et al., 1983). Für die vorliegende Arbeit ist zentral, dass als Sozialisationsinstanz neben den Eltern die Schulklasse mit Lehrkräften, Mitschüler(inne)n und Unterrichtsmedien eine bedeutsame Rolle spielen (Bussey & Bandura, 1999; Dresel et al., 2001; Finsterwald & Ziegler, im Druck). Für den mathematisch-naturwissenschaftlichen Bereich ist beispielsweise gut belegt, dass Lehrkräfte häufig geschlechtsspezifische Erwartungen und Überzeugungen aufweisen, wonach Jungen effektiver lernen würden und generell begabter seien als Mädchen (z. B. Heller, Finsterwald, & Ziegler, 2000; Rustemeyer, 1999). Diese Muster finden ihre Entsprechung in geschlechtsspezifischen Interaktionsstilen zwischen Lehrkräften und Schüler(inne)n (zsf. Jones & Dindia, 2004). Weiterhin können Überzeugungen und Wertesysteme von Mitschüler(inne)n geschlechtsbezogene Selbstbilder noch verstärken bzw. stabilisieren (vgl. Wang, Haertel & Walberg, 1990) und die Interaktion zwischen Mädchen und Jungen im Klassenrahmen spielt eine wichtige Rolle für die Entstehung und Aufrechterhaltung von Motivations-, Lern- und Leistungsunterschieden (Hannover & Kessels, 2002; Hoffmann et al., 1997).

### *Unterschiede zwischen Klassen und Schulen im Ausmaß von Geschlechtsunterschieden*

Auf der Grundlage dieser – mit Lehrkräften, Klassenkamerad(inn)en und Unterrichtsmedien als Sozialisator(inn)en – multiplen Trägerschaft der geschlechtsspezifischen Sozialisation in der Schulklasse und der bei den ein-

zelen Träger(inne)n auftretenden Varianz in der Ausprägung sozialisationsrelevanter Faktoren, nehmen wir an, dass zwischen Schulklassen Unterschiede dahingehend bestehen, wie stark in ihnen geschlechtsspezifische Erwartungen, Überzeugungen und Interaktionsstile zum Tragen kommen. Dementsprechend erwarten wir, dass eine substantielle *Varianz im Ausmaß der Geschlechtsunterschiede zwischen Klassen und Schulen* besteht (analog zu der großen Varianz auf Ebene der Staaten). Im Interpretationsrahmen sozialisationstheoretischer Erklärungsansätze bedeuten diese Schwankungen, dass die Sozialisationserfahrungen von Jungen und Mädchen in unterschiedlichen Klassen und Schulen in unterschiedlich starkem Ausmaß durch differierende Rollenerwartungen und Stereotype charakterisiert sind (vgl. Freeman, 2004). So mag es geschlechtsegalitäre Klassen geben, in denen Geschlechtsunterschiede in Leistung und Motivation nicht oder nur abgeschwächt zum Tragen kommen und die im Kontrast zu besonders geschlechtsdiskriminativen Klassen stehen, in denen Differenzen zwischen den Geschlechtern besonders stark ausgeprägt sind. Diese Varianz zwischen Klassen und Schulen im Ausmaß von Leistungs- und Motivationsunterschieden wurde bislang allerdings kaum untersucht.

Eine Varianz im Ausmaß von Geschlechtsunterschieden könnte dabei nicht nur mit unterschiedlichen Ausprägungen von geschlechtsspezifischen Erwartungen und Überzeugungen bei Lehrkräften und Mitschüler(inne)n sowie unterschiedlich stark vorhandenen Geschlechtsrollenstereotypen in den verwendeten Unterrichtsmedien zusammenhängen, sondern auch mit Merkmalen der Klassenstruktur, insbesondere der *Geschlechterverteilung* und der *Klassengröße*.

Hinsichtlich der Geschlechterverteilung wären zwei gegensätzliche Szenarien denkbar. Einerseits wäre es möglich, dass bei unausgewogenem Geschlechterverhältnis gegenüber einer Minderheit mit schlechteren Leistungen ein Mildeeffekt bei der Leistungsbewertung oder

eine besondere Förderung durch die Lehrkräfte wirksam werden. Danach sollte etwa in Klassen mit hohem Mädchenanteil der Geschlechtsunterschied in sprachlichen Fächern zu Ungunsten der Jungen abgeschwächt sein. Andererseits könnte in solchen Klassen den Leistungen der überzähligen Mädchen eine normative Funktion zukommen und als Referenz für die Bewertung der durchschnittlich schlechteren Leistungen der Jungen fungieren, was zu einer Vergrößerung der Geschlechtsunterschiede führen sollte.

Hinsichtlich der Klassengröße könnten Unterschiede im Ausmaß von Geschlechtsunterschieden auch durch eine in unterschiedlich großen Klassen unterschiedlich starke Aktivierung von geschlechtsbezogenen Erwartungen und Überzeugungen seitens der Lehrkräfte zustande kommen: So kann auf der Grundlage des Befunds, dass Lehrkräfte in größeren Klassen ein höheres Ausmaß an Stress und kognitiver Belastung erleben (Schrader, Helmke, Hosenfeld & Ridder, 2001), und des in der sozialpsychologischen Literatur dokumentierten Effekts, dass der Einfluss von Stereotypen bei hoher kognitiver Belastung besonders stark ist (Sherman, Lee, Bessenoff & Frost, 1998), erwartet werden, dass Lehrkräfte mit zunehmender Klassengröße in stärkerem Ausmaß Geschlechtsrollenstereotype aktivieren. In Folge dessen sollte das Ausmaß von Geschlechtsunterschieden mit zunehmender Klassengröße steigen.

Schließlich könnten Unterschiede zwischen Klassen und Schulen im Ausmaß von Geschlechtsunterschieden unter jahrgangsstufen- und schulformübergreifender Perspektive auch in Form von zwei Schereneffekten zu Tage treten: Erstens sind *temporale Schereneffekte* zu erwarten, wonach die Geschlechterdifferenzen im Laufe der Schulzeit an Substanz gewinnen (Hill & Lynch, 1983; Trautner, Helbing, Sahn & Lohaus, 1988). Dafür spricht etwa, dass sich im Grundschulbereich Geschlechtsunterschiede hinsichtlich der mathematischen Schulleistung nicht und hinsichtlich der mathematikbezogenen Motiva-

tion nur teilweise nachweisen lassen (z. B. Tiedemann & Faber, 1994), während die entsprechenden Geschlechtsunterschiede zu Ungunsten der Mädchen in der Sekundarstufe deutlich ausgeprägt sind (vgl. Kerr, 1994). Diese Zunahme an Unterschieden zwischen Jungen und Mädchen kann dabei sowohl als zunehmender Niederschlag von geschlechtsspezifischen Sozialisierungseinflüssen verstanden als auch auf die Entwicklungsaufgabe der Ausgestaltung der eigenen Geschlechtsidentität zurückgeführt werden (Hannover, 2004). Zweitens sind auch *vertikale Schereneffekte* zu vermuten, wonach das Ausmaß von Geschlechterdifferenzen dahingehend schulformspezifisch ausfällt, dass sie in höheren Bildungsgängen abgeschwächt und in niedrigeren Bildungsgängen verstärkt sind. Untermauert wird dies etwa durch das Ergebnis von PISA 2003, wonach die Lesekompetenz von Jungen auf den untersten Kompetenzstufen besonders deutlich unter jener der Mädchen liegt (Zimmer et al., 2004). Folglich ist zu erwarten, dass die Geschlechtsunterschiede im sprachlichen Bereich in der Hauptschule (und abgeschwächt: in der Realschule) stärker ausgeprägt sind als am Gymnasium. Unter sozialisationstheoretischer Perspektive lassen vertikale Schereneffekte auf eine mit höherer Schulform zunehmend schwächere Ausprägung von geschlechtsspezifischen Erwartungen, Überzeugungen und Interaktionsstilen schließen, die möglicherweise auf Zusammenhänge mit dem Sozialstatus sowie auf unterschiedliche Beteiligungsraten von Mädchen und Jungen an den einzelnen Schulformen zurückzuführen sein könnten (vgl. Hosenfeld, Köller & Baumert, 1999; Stanat & Kunter, 2001, 2003).

### *Fragestellungen und Hypothesen*

Unsere zentrale Fragestellung richtete sich darauf, ob sich Schulklassen und Schulen im Ausmaß von Geschlechtsunterschieden bei Schulleistung und Motivation unterscheiden. Um differenzierte Aussagen ableiten zu können, fokussierten wir mit der kompletten Se-

kundarstufe einen breiten Jahrgangsstufenbereich, mit Hauptschule, Realschule und Gymnasium alle drei Schulformen des Regelschulsystems und mit Mathematik, Deutsch und der ersten Fremdsprache Fächer sowohl aus dem mathematisch-naturwissenschaftlichen als auch dem sprachlichen Bereich. Als Indikator der Schulleistung bezogen wir Leistungsbewertungen durch Lehrkräfte (Zensuren) ein. Um erste Anhaltspunkte zu Klassen- und Schuleffekten bei Geschlechtsunterschieden im motivationalen Bereich zu erhalten, betrachteten wir als klassischen Indikator Leistungsaspirationen, im Sinne der Diskrepanz zwischen vorangehendem Ergebnis (vorangehend erhaltene Zensur) und Ansprüchen an nachfolgende Leistungen (gerade noch zufriedenstellende Zensur), die sog. Zieldiskrepanz (Heckhausen, 1989; Lewin, Dembo, Festinger & Sears, 1944). Dieser Indikator kann theoretisch als makroskopisches Resultat verschiedener motivationaler (Mikro-)Prozesse aufgefasst werden, die u. a. die Suche nach Ursachen für vorangegangene Leistungen, die Bewertung der Aufgabenstellung sowie die Abschätzung der Erfolgserwartung auf der Grundlage der Aufgabenschwierigkeit und der Einschätzung der eigenen Fähigkeiten umfassen (vgl. Heckhausen, 1989; Pintrich & Schunk, 2002).

Unsere Hypothesen gliedern sich in drei aufeinander aufbauende Gruppen. Die beiden Basishypothesen der ersten Gruppe betreffen die bereits dokumentierten Geschlechterdifferenzen in Mathematik und den beiden sprachlichen Fächern. Da Geschlechtsunterschiede durch Jahrgangsstufen- und Schulformeffekte überformt oder verdeckt sein können (vgl. Zimmer et al., 2004), wurde deren Kontrolle explizit berücksichtigt:

H<sub>1a</sub>: Nach Kontrolle von Jahrgangsstufe und Schulform bestehen durchschnittliche Geschlechtsunterschiede dahingehend, dass Mädchen in Mathematik schlechtere Leistungsbewertungen erhalten und geringere Leistungsaspirationen aufweisen als Jungen.

**H<sub>1b</sub>:** Nach Kontrolle von Jahrgangsstufe und Schulform bestehen durchschnittliche Geschlechtsunterschiede dahingehend, dass Mädchen in Deutsch und der ersten Fremdsprache bessere Leistungsbewertungen erhalten und höhere Leistungsaspirationen aufweisen als Jungen.

Aufbauend auf der ausgeführten Argumentation für Unterschiede in der Ausprägung geschlechtsspezifischer Erwartungen, Überzeugungen und Interaktionsstile zwischen Schulklassen und Schulen thematisieren die beiden zentralen Hypothesen der zweiten Gruppe Klassen- und Schulunterschiede im Ausmaß der Geschlechterdifferenzen bei Leistungsbewertungen und Leistungsaspirationen:

**H<sub>2a</sub>:** Schulklassen unterscheiden sich dahingehend, wie stark in ihnen Geschlechtsunterschiede bei Leistungsbewertungen und Leistungsaspirationen ausgeprägt sind.

**H<sub>2b</sub>:** Schulen unterscheiden sich dahingehend, wie stark in ihnen Geschlechtsunterschiede bei Leistungsbewertungen und Leistungsaspirationen ausgeprägt sind.

Schließlich fassen die folgenden vier Hypothesen der dritten Gruppe unsere dargelegten Überlegungen und darauf basierenden Erwartungen zum Einfluss struktureller Merkmale von Klassen und Schulen auf das Ausmaß der erwarteten Geschlechtsunterschiede zusammen:

**H<sub>3a</sub>:** Die Stärke von Geschlechtsunterschieden bei Leistungsbewertungen und Leistungsaspirationen hängt mit der Geschlechterverteilung in der Klasse zusammen.

**H<sub>3b</sub>:** Die Stärke von Geschlechtsunterschieden bei Leistungsbewertungen und Leistungsaspirationen nimmt mit zunehmender Klassengröße zu.

**H<sub>3c</sub>:** Die Stärke von Geschlechtsunterschieden bei Leistungsbewertungen und Leistungsaspirationen nimmt mit ansteigender Jahrgangsstufe zu (temporaler Schereneffekt).

**H<sub>3d</sub>:** In der Hauptschule und der Realschule bestehen stärkere Geschlechtsunterschiede bei Leistungsbewertungen und Leistungsaspirationen als am Gymnasium (vertikaler Schereneffekt).

## Methoden

Es wurde ein Teil der Stichprobe analysiert, die in den Jahren 2003 bis 2005 für die Normierung der von den Autor(innen) entwickelten Ulmer Motivationstestbatterie rekrutiert wurde (*UMTB*; Ziegler, Dresel, Schober & Stöger, in Vorb.). Einbezogen wurden die Jahrgangsstufen 5 bis 9 an Hauptschulen sowie 5 bis 10 an Realschulen und Gymnasien. Gemäß der formulierten Hypothesen wurden als abhängige Variablen Leistungsbewertungen (Zensuren) sowie Leistungsaspirationen in drei Hauptfächern (Mathematik, Deutsch, erste Fremdsprache) und als unabhängige Variablen biografische Angaben sowie strukturelle Merkmale der Klassen und Schulen (Geschlecht, Jahrgangsstufe, Geschlechterverteilung in der Klasse, Klassengröße, Schulform) einbezogen. Die Geschlechtsunterschiede wurden hierarchisch linear auf drei Ebenen modelliert (*Individual-, Klassen- und Schulebene*).

## Stichprobe

Die Stichprobe umfasste 9118 Schüler (4438 Jungen und 4680 Mädchen) aus 492 Klassen an 54 koedukativen Regelschulen in Baden-Württemberg. Eine Aufgliederung der Stichprobe nach Jahrgangsstufe und den drei einbezogenen Schulformen findet sich in Tabelle 1. Die Klassengröße differenzierte zwischen 11 und 33 Schüler(inne)n ( $M = 26.5$ ;  $SD = 4.5$ ), die Anzahl der Klassen pro Schule lag in unserer Stichprobe zwischen 1 und 29 ( $M = 9.1$ ;  $SD = 6.6$ ). Die Geschlechterverteilung in den Klassen variierte zwischen 8 % und 94 % Mädchenanteil ( $M = 51$  %;  $SD = 15$  %). Das Durchschnittsalter betrug 13.6 Jahre ( $SD = 1.76$ ). Die überwiegende Zahl der Schüler(innen) hatte Englisch (96 %), lediglich 3 % hatte Latein und 1 % Französisch als erste Fremdsprache.

## Erfasste Maße

Die *Leistungsbewertungen* in den Fächern Mathematik und Deutsch wurden mit je einem Item erfasst („Welche Mathematik-/Deutsch-Note hattest du im letzten Zeugnis?“). Zur Erfassung der Leistungsbewertung in der ersten Fremdsprache wurde diese zunächst erfragt („Welches ist die Fremdsprache, die du seit der 5. Klasse hast?“), und anschließend analog zu den anderen beiden Hauptfächern die Zensur abgefragt („Welche Note hattest du in dieser Fremdsprache im letzten Zeugnis?“). Schüler(innen) der 5. Jahrgangsstufe wurden um die Angabe der Zensuren in der letzten Klassenarbeit gebeten, da das vorangegangene Zeugnis häufig das Jahreszeugnis der 4. Jahrgangsstufe war und damit in den wenigsten Fällen eine Fremdsprachennote aufwies. Vor dem Hintergrund einschlägiger Arbeiten (z. B. Dickhäuser & Plenter, im Druck) kann davon ausgegangen werden, dass die Übereinstimmung zwischen selbst berichteten und tatsächlich erhaltenen Zensuren

Tabelle 1: Anzahl Klassen (K) und Schüler(innen) (S) in der Stichprobe, aufgegliedert nach Jahrgangsstufe und Schulform

	Hauptschulen (27 Schulen)		Realschulen (14 Schulen)		Gymnasien (13 Schulen)		Gesamt (54 Schulen)	
	K	S	K	S	K	S	K	S
5. Jahrgangsstufe	27	409	29	564	39	826	95	1799
6. Jahrgangsstufe	30	439	26	563	38	786	94	1788
7. Jahrgangsstufe	35	577	24	495	34	669	93	1741
8. Jahrgangsstufe	34	510	26	484	28	555	88	1549
9. Jahrgangsstufe	30	485	24	506	28	524	82	1515
10. Jahrgangsstufe			12	235	28	491	40	726
Gesamt	156	2420	141	2847	195	3851	492	9118

außerordentlich hoch ist und dass – im Zusammenhang der vorliegenden Arbeit besonders wichtig – bei den Selbstberichten keine geschlechterdifferenziellen Verzerrungen auftreten.

Als Maß für *Leistungsaspirationen* wurde die schlechteste, noch zufriedenstellende Zensur in Klassenarbeiten erfasst („Mit welcher Note in Mathematik-/Deutsch-Klassenarbeiten bist du gerade noch zufrieden?“, „Mit welcher Note in Klassenarbeiten bist du in dieser Fremdsprache gerade noch zufrieden?“). In allen Analysen mit Leistungsaspirationen als abhängige Variable wurden Einflüsse der tatsächlich erzielten Leistungen durch Aufnahme der korrespondierenden Zensur als unabhängige Variable auspartialisiert. Damit spiegeln die resultierenden Werte die um Einflüsse der Leistungsgüte bereinigte Zieldiskrepanz zwischen vorangegangener Leistung und Ansprüchen an nachfolgende Leistungen wider (vgl. Heckhausen, 1989; Lewin et al., 1944).

### Analysemethodik

Die in unseren Hypothesen formulierten Aussagen betreffen durchschnittliche Unterschiede zwischen Jungen und Mädchen, Unterschiede zwischen Schulklassen bzw. Schulen im Ausmaß dieser Geschlechtsunterschiede sowie Merkmale von Schulklassen bzw. Schulen als Prädiktoren des klassen- bzw. schulspezifischen Ausmaßes der Geschlechtsunterschiede. Damit werden simultan drei ineinander verschachtelte Ebenen betrachtet, auf denen Prädiktoren und Kriterien angesiedelt sind: Schüler(innen) in Klassen an Schulen. Die angemessene methodische Repräsentation der Effekte von individuellen, klassenspezifischen und schulspezifischen Prädiktoren auf Leistungsbewertungen und Leistungsaspirationen erfordert eine hierarchisch lineare Modellierung als Dreiebenenmodell. Nur damit kann der verschachtelten Datenstruktur und möglichen Wechselwirkungen zwischen Prädik-

toren auf unterschiedlichen Ebenen (sog. ebenenübergreifende Wechselwirkungen) Rechnung getragen werden (zur Methodik siehe Snijders & Bosker, 1999).

Mit dem Dreiebenenansatz ist es möglich, den Einfluss einer unabhängigen Variablen – hier: des Geschlechts der Schüler(innen) – auf eine abhängige Variable – hier: Leistungsbewertung oder Leistungsaspiration – getrennt für alle einbezogenen Schulklassen und Schulen zu schätzen. Zur Prüfung der Hypothesen  $H_{1a}$  und  $H_{1b}$  ist der Mittelwert der resultierenden Regressionsgewichte der Geschlechtsvariable bedeutsam; er spezifiziert den durchschnittlichen Geschlechtsunterschied auf der jeweiligen abhängigen Variable. Die Varianz dieser Regressionsgewichte zwischen Klassen bzw. Schulen quantifiziert die für die Prüfung der Hypothesen  $H_{2a}$  und  $H_{2b}$  bedeutsame Unterschiedlichkeit im Ausmaß klassen- bzw. schulspezifischer Geschlechtsunterschiede. Simultan können mit dem Dreiebenenmodell Jahrgangsstufen- und Schulformeffekte auf das vom Geschlecht der Schüler(innen) unabhängige Niveau von Leistungsbewertungen und Leistungsaspirationen kontrolliert werden.

Gemäß dieser Überlegungen wurde zur Prüfung der Hypothesen  $H_{1a}$  bis  $H_{2b}$  ein *Modell 1* spezifiziert, das durch die folgenden Gleichungen definiert ist:

$$\text{Für Leistungsbewertungen:} \\ Y_{ijk} = \pi_{0jk} + \pi_{1jk} \cdot \text{Weiblich} + e_{ijk} \quad (1)$$

$$\text{Für Leistungsaspirationen:} \\ Y_{ijk} = \pi_{0jk} + \pi_{1jk} \cdot \text{Weiblich} + \pi_2 \cdot \text{Zensur} + e_{ijk} \quad (1')$$

$$\text{Für Leistungsbewertungen und -aspirationen:} \\ \pi_{0jk} = \beta_{00k} + \beta_{01} \cdot \text{JGS} + r_{0jk} \quad (2)$$

$$\pi_{1jk} = \beta_{10k} + r_{1jk} \quad (3)$$

$$\beta_{00k} = \gamma_{000} + \gamma_{001} \cdot \text{HS} + \gamma_{002} \cdot \text{RS} + u_{00k} \quad (4)$$

$$\beta_{10k} = \gamma_{100} + u_{10k} \quad (5)$$

In diesem Modell wird die abhängige Variable  $Y_{ijk}$  des/der  $i$ -ten Schülers/Schülerin in der  $j$ -ten Klasse an der  $k$ -ten Schule in Form einer Regressionsgleichung auf Individualebene ausgedrückt, die sein/ihr Geschlecht (Variable *Weiblich*; männlich = -0.5; weiblich = 0.5) als Prädiktor enthält (Gleichung 1). Im Unterschied zu herkömmlichen Regressionsmodellen mit nur einer Analyseebene werden das Regressionsgewicht  $\pi_{1jk}$  der Geschlechtsvariable sowie der Achsenabschnittsparameter  $\pi_{0jk}$  für jede Schulklasse separat geschätzt. In der Folge repräsentieren die Residuen  $e_{ijk}$  ausschließlich Abweichungen innerhalb der jeweiligen Klasse. Wie oben ausgeführt, wurden bei der Analyse der Leistungsaspirationen Einflüsse der Leistungsbewertungen auspartialisiert. In Gleichung 1' wurde dazu zusätzlich die Leistungsbewertung (*Zensur*; zentriert am Stichprobenmittelwert) im jeweiligen Schulfach als Prädiktor eingesetzt, wobei hier keine Varianz des Regressionsgewichts  $\pi_2$  zwischen Klassen oder Schulen zugelassen wurde.

Sowohl der Achsenabschnittsparameter  $\pi_{0jk}$  als auch das Regressionsgewicht  $\pi_{1jk}$  der Geschlechtsvariable wurden als abhängige Variable einer Regressionsgleichung auf Klassenebene spezifiziert (Gleichungen 2 und 3), deren Koeffizienten  $\beta_{00k}$  und  $\beta_{10k}$  in jeder Schule separat geschätzt wurden. Zur Kontrolle von Jahrgangsstufeneffekten auf das (geschlechtsunabhängige) Niveau der abhängigen Variablen wurde in die Gleichung des Achsenabschnittsparameters  $\beta_{0jk}$  die Jahrgangsstufe mit dem Gewicht  $\beta_{01}$  eingesetzt (*JGS*; fixiert, d. h. keine Varianz des entsprechenden Regressionsgewichts zwischen Schulen zugelassen). Die klassenspezifischen Residuen  $r_{0jk}$  und  $r_{1jk}$  repräsentieren ausschließlich Abweichungen von dem für die jeweilige Schule erwarteten Wert und somit ausschließlich Abweichungen zwischen Klassen, aber nicht zwischen Schulen.

In analoger Weise wurden die beiden schulspezifischen Achsenabschnittsparameter  $\beta_{00k}$  und  $\beta_{10k}$  als abhängige Variable je einer Regressionsgleichung auf Schulebene spezifiziert, deren Parameter für die Gesamtstichprobe der Schulen geschätzt wurden (Gleichungen 4 und 5). Zur Kontrolle von Schulformeffekten auf das Niveau der abhängigen Variablen wurden zwei Dummyvariablen *HS* (Hauptschule = 0.5; Sonst = -0.5) sowie *RS* (Realschule = 0.5; Sonst = -0.5) gebildet und mit den Gewichten  $\gamma_{001}$  sowie  $\gamma_{002}$  in Gleichung 4 eingesetzt. Die schulspezifischen Residuen  $u_{00k}$  und  $u_{10k}$  verweisen auf Residualvarianzen zwischen Schulen, die nicht auf Unterschiede zwischen Klassen oder Individuen innerhalb von Schulen zurückgehen.

Nach Zentrierung aller Prädiktoren am Stichprobenmittelwert spiegelt der Parameter  $\gamma_{000}$  die durchschnittliche (und für ungleiche Zellbesetzungen korrigierte) Leistungsbewertung bzw. Leistungsaspiration in der Gesamtstichprobe wider (vgl. Snijders & Bosker, 1999). Die Hypothesen  $H_{1a}$  und

$H_{1b}$  werden anhand des Parameters  $\gamma_{100}$  geprüft, der die durchschnittlichen Geschlechtsunterschiede nach Kontrolle von Jahrgangsstufen- und Schulformeffekten quantifiziert. Schließlich repräsentieren die zu Kontrollzwecken in Modell 1 aufgenommenen und für die verfolgte Fragestellung nicht zentralen Parameter  $\beta_{01}$ ,  $\beta_{001}$  und  $\beta_{002}$  die geschlechtsunabhängigen Effekte der Jahrgangsstufe und der Schulform.

Ein wesentlicher Vorteil des Dreiebenenmodells liegt in der Dekomposition der Residualvarianz auf die drei Ebenen. Dabei entspricht  $E = \text{Var}(e_{ijk})$  der Residualvarianz der abhängigen Variablen innerhalb von Klassen,  $R_0 = \text{Var}(r_{0jk})$  jener zwischen Klassen und  $U_{00} = \text{Var}(u_{00k})$  jener zwischen Schulen. Mit der Schätzung des „leeren Modells“ ohne Prädiktoren (definiert durch die Gleichungen  $Y_{ijk} = p_{0jk} + e_{ijk}$ ,  $p_{0jk} = \beta_{00k} + r_{0jk}$  und  $\beta_{00k} = \gamma_{000} + u_{00k}$ ) können so die Anteile der Gesamtvarianz einer abhängigen Variable bestimmt werden, die innerhalb von Klassen ( $E$ ), zwischen Klassen ( $R_0$ ) und zwischen Schulen ( $U_{00}$ ) auftreten. Für die vorliegende Fragestellung sind die in der regulären Kriteriumsvarianz nicht enthaltenen Varianzanteile  $R_1 = \text{Var}(r_{1jk})$  und  $U_{10} = \text{Var}(u_{10k})$  zentral. Sie schätzen die Varianz des Parameters  $\beta_{10k}$  zwischen Klassen sowie die Varianz des Parameters  $\gamma_{100}$  zwischen Schulen. Damit geben sie an, wie stark das Ausmaß von Geschlechtsunterschieden zwischen Schulklassen und zwischen Schulen variiert. Folglich sind sie die beiden relevanten Kenngrößen zur Prüfung der zentralen Hypothesen  $H_{2a}$  und  $H_{2b}$ .

Die mit den Hypothesen  $H_{3a}$  bis  $H_{3d}$  erwarteten Einflüsse der Geschlechterverteilung, der Klassengröße, der Jahrgangsstufe und der Schulform auf das Ausmaß von Geschlechtsunterschieden werden im Dreiebenenansatz als Kovariation dieser Strukturmerkmale mit dem auf Individualebene angesiedelten Einfluss des Geschlechts der Schüler(innen) auf die abhängigen Variablen modelliert. Dazu wurden die genannten Prädiktoren in die beiden in Modell 1 enthaltenen Gleichungen 3 und 5 für das klassen- bzw. schulspezifische Regressionsgewicht der Geschlechtsvariable eingesetzt. Das resultierende Modell 2 ist ein *Slopes-as-Outcomes-Model*, in dem der Steigungsparameter der Variable *Weiblich* durch unabhängige Variablen auf den höheren Analyseebenen prädiert wird. Im Falle einer statistisch bedeutsamen Prädiktion liegt eine ebenenübergreifende Interaktion vor, bei welcher der Einfluss der Geschlechtsvariable auf Individualebene durch eine Variable auf Klassen- oder Schulebene moderiert wird (vgl. Snijders & Bosker, 1999).

Die gegenüber Modell 1 modifizierten Gleichungen von Modell 2 lauten:

$$\pi_{1jk} = \beta_{10k} + \beta_{11} \cdot \text{AntWeiblich} + \beta_{12} \cdot \text{KGröße} + \beta_{13} \cdot \text{JGS} + r_{1jk} \quad (3')$$

$$\beta_{10k} = \gamma_{100} + \gamma_{101} \cdot \text{HS} + \gamma_{102} \cdot \text{RS} + u_{10k} \quad (5')$$

In Gleichung 3' wurden als Prädiktoren des klassenspezifischen Regressionsgewichts  $\pi_{ijk}$  der Geschlechtsvariable die folgenden Strukturmerkmale von Schulklassen eingesetzt: die Geschlechterverteilung in der Klasse als Mädchenanteil (*AntWeiblich*; 0.00 – 1.00), die Klassengröße als Anzahl der Schüler(innen) (*KGröße*) und die Jahrgangsstufe (*JGS*). Die Effekte dieser Variablen wurden fixiert, Variationen der Parameter  $\beta_{11}$ ,  $\beta_{12}$  und  $\beta_{13}$  zwischen Schulen wurden also nicht zugelassen. In Gleichung 5' wurde die Schulform (*HS* und *RS*) als Prädiktor des schulspezifischen Durchschnittswerts  $\beta_{10k}$  des Regressionsgewichts  $\pi_{ijk}$  eingesetzt. Um mittlere Werte der Moderatoreffekte für die Gesamtstichprobe zu erhalten, wurden alle Prädiktoren wiederum am Stichprobenmittelwert zentriert.

Um die Interpretation der Regressionskoeffizienten zu erleichtern, wurden die abhängigen Variablen vor der Analyse aller Modelle z-standardisiert. Auf diese Weise können die Regressionskoeffizienten wie standardisierte Mittelwertdifferenzen (Effektgröße  $d = (M_1 - M_2)/s$  mit  $s$  als Standardabweichung in der Gesamtstichprobe) zwischen zwei Prädiktorausprägungen interpretiert werden, die den Abstand von einer Skaleneinheit aufweisen. Beispielsweise spiegelt der Koeffizient  $\gamma_{100}$  der Variable *Weiblich* (männlich = -0.5; weiblich = 0.5; Abstand = 1) die standardisierte Differenz zwischen dem Stichprobenmittelwert der Mädchen und jenem der Jungen und damit die Effektgröße des durchschnittlichen Geschlechtsunterschieds wider (analog: Differenz zwischen reinen Jungen- und reinen Mädchenklassen, Differenz zwischen zwei Klassen mit um eine Person unterschiedlichen Größen, Differenz zwischen zwei benachbarten Jahrgangsstufen, Differenz zwischen zwei Schulformen). Zur zusätzlichen Erleichterung wurden die abhängigen Variablen umgepolt, so dass ein hoher Wert einer guten Leistungsbewertung bzw. einer hohen Anspruchsniveausetzung entspricht.

Die Analysen wurden mit *HLM5* durchgeführt (Raudenbush, Bryk, Cheong & Congdon, 2001). Verbesserungen in der Modellanpassung wurden mit Devianztests geprüft (vgl. Snijders & Bosker, 1999, S. 88 – 91).

## Ergebnisse

### Deskriptive Statistiken

Tabelle 2 enthält Mittelwerte und Standardabweichungen für die abhängigen Variablen. Die durchschnittlichen Zensuren in den drei Hauptfächern waren sehr ähnlich und lagen knapp oberhalb der Note „Befriedigend“ (3). Die berichteten Leistungsaspirationen, also die Zensuren, mit denen die Schüler(innen) gerade noch zufrieden waren, lagen auffällig nah an den tatsächlichen Leistungsbewertun-

gen. Dies spricht dafür, dass die erzielten Leistungen im Mittel gerade noch als zufriedenstellend angesehen wurden.

Tabelle 2: Mittelwerte und Standardabweichungen von Leistungsbewertungen und Leistungsaspirationen in Mathematik, Deutsch und der ersten Fremdsprache

	<i>M</i>	<i>SD</i>
<i>Mathematik</i>		
Leistungsbewertung	2.79	0.94
Leistungsaspiration	2.96	0.68
<i>Deutsch</i>		
Leistungsbewertung	2.84	0.87
Leistungsaspiration	3.08	0.57
<i>Erste Fremdsprache</i>		
Leistungsbewertung	2.81	0.92
Leistungsaspiration	3.00	0.66

Tabelle 3 enthält die Ergebnisse zu den „leeren Modellen“. Wie oben ausgeführt, wurden bei den Leistungsaspirationen – wie in allen nachfolgenden Analysen – die erwartungsgemäß moderat positiven Abhängigkeiten von Unterschieden in den tatsächlichen Leistungsbewertungen auspartialisiert. Die Schätzung der Varianzanteile erbrachte, dass zwischen 6 % und 13 % der Varianz der tatsächlichen Zensuren und zwischen 4 % und 6 % der Varianz der gerade noch zufriedenstellenden Zensuren auf Unterschiede zwischen Schulen zurückgehen. Weitere 8 % der Varianz der Leistungsbewertungen und 4 % bis 5 % der Varianz der Leistungsaspirationen repräsentieren Unterschiede zwischen Klassen innerhalb von Schulen. Da die Varianz innerhalb von Klassen typischerweise weit größer ist als die Varianz zwischen Klassen und Schulen – Snijders und Bosker (1999) nennen 5 % bis 20 % als typische Werte für die auf den beiden höheren Ebenen insgesamt angesiedelte Varianz –, liegt die Größe der beobachteten Varianzanteile im erwarteten Rahmen. Alle genannten Varianzen zwischen Schulen und Klassen waren signifikant von Null verschieden, so dass bei allen abhängigen Variablen statistisch bedeutsame Klassen- und Schuleffekte konstatiert werden können.

Tabelle 3: Schätzung der „leeren Modelle“ zur Bestimmung der ebenenspezifischen Varianzanteile von Leistungsbewertungen und Leistungsaspirationen

	Mathematik		Deutsch		Erste Fremdsprache	
	Leistungs- bewertung	Leistungs- aspiration	Leistungs- bewertung	Leistungs- aspiration	Leistungs- bewertung	Leistungs- aspiration
<i>Regressionskoeffizienten</i>						
Achsenabschnitt ( $\gamma_{000}$ )	-0.06	0.08	-0.12	0.07	-0.04	0.07
Leistungsbewertung ( $\pi_2$ )		0.54***		0.38***		0.55***
<i>Varianzanteile</i>						
Residualvarianz innerhalb Klassen ( $E$ )	0.86	0.67	0.79	0.80	0.86	0.66
Residualvarianz zwischen Klassen ( $R_0$ )	0.08***	0.03***	0.08***	0.05***	0.08***	0.03***
Residualvarianz zwischen Schulen ( $U_{00}$ )	0.06***	0.05***	0.13***	0.04***	0.06***	0.03***
<i>Modellanpassung</i>						
Devianz	25150.1	22651.7	24476.1	24293.4	25111.4	22415.9

Anmerkungen: Alle Variablen wurden z-standardisiert und so gepolt, dass ein hoher Wert eine gute Leistungsbewertung bzw. eine hohe Leistungsaspiration widerspiegelt.

\*\*\*  $p < .001$

### *Durchschnittliche Geschlechtsunterschiede nach Kontrolle von Schulform und Jahrgangsstufe*

Tabelle 4 liefert einen Überblick der Ergebnisse zu Modell 1, mit dem die durchschnittlichen Geschlechtsunterschiede in den drei Fächern nach Kontrolle von Jahrgangsstufen- und Schulformeffekten bestimmt wurden. Für alle sechs abhängigen Variablen lieferte dieses Modell eine bessere Datenanpassung als das „leere Modell“ ( $\chi^2(12) > 84.8$ ;  $p < .001$ ). Die Residualvarianzen wurden gegenüber diesem vor allem auf Schulebene erheblich reduziert, was darauf hindeutet, dass Schulunterschiede bei Leistungsbewertungen und Leistungsaspirationen vor allem auf Schulformunterschiede zurückzuführen sind (vgl. Hosenfeld, Helmke, Ridder & Schrader, 2001).

Im Fach Mathematik waren die erwarteten Geschlechtsunterschiede zu beobachten: Jungen erhielten im Mittel bessere Leistungsbewertungen und verfolgten höhere Leistungsan-

sprüche als Mädchen. Der Effekt bei den – für die Motivation der Schüler(innen) indikativen – gerade noch zufriedenstellenden Zensuren (standardisierte Mittelwertdifferenz  $d = \gamma_{100} = -0.25$ ) fiel dabei erkennbar größer aus als der Effekt bei den tatsächlich erhaltenen Zensuren ( $d = -0.15$ ).

In den beiden sprachlichen Fächern erhielten Mädchen erwartungsgemäß die durchschnittlich besseren Leistungsbewertungen: Besonders groß fiel der Geschlechtsunterschied im Fach Deutsch mit einer Effektgröße von  $d = 0.42$  aus, in der ersten Fremdsprache lag er bei  $d = 0.17$ . Bei Betrachtung der Leistungsaspirationen im sprachlichen Bereich konnten wir für das Fach Deutsch eine höhere Anspruchsniveausetzung von Mädchen statistisch absichern, wobei der Effekt aber deutlich kleiner ausfiel als bei den korrespondierenden Leistungsbewertungen ( $d = 0.13$ ). Für die erste Fremdsprache war kein Geschlechtseffekt bei den Leistungsaspirationen statistisch abzusichern ( $d = -0.01$ ).

Tabelle 4: Schätzung von Modell 1 zur Bestimmung der durchschnittlichen Geschlechtsunterschiede bei Leistungsbewertungen und Leistungsaspirationen unter Kontrolle von Jahrgangsstufe und Schulform sowie zur Quantifizierung der Unterschiede zwischen Klassen und Schulen im Ausmaß der Geschlechtsunterschiede

	Mathematik		Deutsch		Erste Fremdsprache	
	Leistungs- bewertung	Leistungs- aspiration	Leistungs- bewertung	Leistungs- aspiration	Leistungs- bewertung	Leistungs- aspiration
<i>Regressionskoeffizienten</i>						
Achsenabschnitt ( $\gamma_{000}$ )	-0.11	0.09	-0.13	0.08	-0.07	0.08
Weiblich ( $\gamma_{100}$ )	-0.15***	-0.25***	0.42***	0.13***	0.17***	-0.01
Jahrgangsstufe ( $\beta_{01}$ )	-0.09***	-0.04***	0.01	-0.05***	-0.09***	-0.04***
Hauptschule ( $\gamma_{001}$ )	-0.52***	0.40***	-0.74***	0.31***	-0.43***	0.35***
Realschule ( $\gamma_{002}$ )	-0.33***	0.09**	-0.45***	0.01	-0.34***	0.04
Leistungsbewertung ( $\pi_2$ )		0.52***		0.37***		0.55***
<i>Varianzanteile</i>						
Residualvarianz innerhalb Klassen ( $E$ )	0.85	0.65	0.74	0.79	0.84	0.65
Residualvarianz zwischen Klassen ( $R_0$ )	0.05***	0.03***	0.08***	0.04***	0.06***	0.03***
Residualvarianz zwischen Schulen ( $U_{00}$ )	0.01***	0.01***	0.03***	0.01***	0.02***	0.01***
Gewicht Weiblich zwischen Klassen ( $R_1$ )	0.04***	0.03***	0.08***	0.04***	0.06***	0.03***
Gewicht Weiblich zwischen Schulen ( $U_{00}$ )	0.00	0.00	0.00	0.00	0.00	0.00
<i>Modellanpassung</i>						
Devianz	24945.4	22344.7	23884.6	24177.0	24911.6	22331.0

Anmerkungen: Alle abhängigen Variablen wurden z-standardisiert und so gepolt, dass ein hoher Wert eine gute Leistungsbewertung bzw. eine hohe Leistungsaspiration widerspiegelt. Alle Prädiktoren wurden am Stichprobenmittelwert zentriert.

\*\*\*  $p < .001$  \*\*  $p < .01$

### *Klassenunterschiede im Ausmaß von Geschlechtsunterschieden*

Im Hinblick auf Klassenunterschiede im Ausmaß der dargestellten Geschlechtsunterschiede waren für alle sechs abhängigen Variablen signifikant von Null verschiedene Varianzen des Regressionsgewichts der Geschlechtsvariable auf Klassenebene (Varianzanteil  $R_1$ ) zu beobachten (vgl. Tabelle 4). Diese weisen darauf, dass sich *Schulklassen* stärker als es aufgrund von Zufallseinflüssen (z. B. hinsichtlich unterschiedlicher Begabungen ihrer Mitglieder) zu erwarten wäre, darin unterschieden, wie stark in ihnen Geschlechts-

unterschiede bei Leistungsbewertungen und Leistungsaspirationen ausgeprägt sind. Für die Anspruchsniveausetzung in der ersten Fremdsprache galt dies unabhängig davon, dass hier kein signifikanter durchschnittlicher Geschlechtseffekt nachgewiesen werden konnte.

Zur weiteren Aufklärung der Klasseneffekte im Ausmaß von Geschlechtsunterschieden wurden die klassenspezifischen Ausprägungen der Geschlechtsunterschiede bei den sechs abhängigen Variablen einer eingehenderen Analyse unterzogen. Dazu wurden für die Verteilungen der klassenspezi-

fischen Regressionsgewichte  $\pi_{1jk}$ , die als standardisierte Mittelwertdifferenzen  $d$  zwischen den Werten von Mädchen und jenen von Jungen innerhalb der betreffenden Schulklasse interpretiert werden können, verschiedene Perzentile berechnet. In Abbildung 1 sind für die Stichprobe der 492 einbezogenen Schulklassen 5 %-, 25 %-, 50 %-, 75 %- und 95 %-Perzentil sowie Minimum und Maximum dargestellt. Für alle drei Fächer verweisen die Verteilungen sowohl auf Klassen, in denen Geschlechtsunterschiede stark ausgeprägt sind, als auch auf Klassen, in denen die Geschlechtsvariable eine sehr geringe oder keine Bedeutung hat.

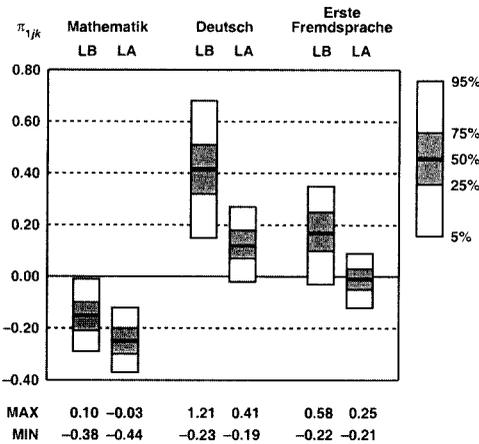


Abbildung 1: Klassenunterschiede im Ausmaß von Geschlechterdiskrepanzen bei Leistungsbewertungen (LB) und Leistungsaspirationen (LA): Perzentilbänder der klassenspezifischen Mittelwertdifferenzen  $\pi_{1jk}$  zwischen Mädchen und Jungen in 492 Schulklassen (in Standardabweichungen der betreffenden abhängigen Variable, positive  $\pi_{1jk}$  stellen günstigere Werte der Mädchen dar)

In Bezug auf das Fach Mathematik verweisen die Perzentile auf kleine bis moderate Unterschiede zwischen Klassen. In der Klasse, in der das (negative) Regressionsgewicht  $\pi_{1jk}$  der Geschlechtsvariable bei der Prädiktion der Mathematikzensuren größer als in 5 % aller Klassen war, entsprach dieses einer standardisierten Mittelwertdifferenz von  $d = -0.29$

und damit einem gegenüber dem durchschnittlichen Geschlechtsunterschied um  $d = 0.14$  zu Ungunsten der Mädchen verstärkten Geschlechtseffekt. Dagegen waren in jener Klasse, in der das Regressionsgewicht größer als in 95 % aller Klassen war, mit  $d = -0.01$  praktisch keine Unterschiede zwischen den Mathematikzensuren der Mädchen und Jungen vorhanden. Der Unterschied zwischen diesen beiden Klassen (entspricht der Perzentilbandlänge) beträgt also etwas mehr als eine viertel Standardabweichung. Darüber hinaus erhielten Mädchen – im Gegensatz zum mittleren in der Gesamtstichprobe beobachteten Geschlechtseffekt – in 4 % der Klassen bessere Mathematikzensuren als Jungen. Hinsichtlich der Anspruchsniveausetzung im Fach Mathematik waren ähnlich große Unterschiede zwischen Schulklassen zu beobachten (Perzentilbandlänge:  $d = 0.25$ ). In 24 % der Schulklassen verfolgten die Mädchen Leistungsansprüche, die  $d = 0.30$  oder mehr unter jenen der Jungen lagen. Andererseits waren in 10 % der Schulklassen diese Unterschiede mit  $d = 0.10$  oder weniger nur sehr schwach ausgeprägt.

Die größten Klasseneffekte zeigten sich bei der Deutschnote. In der Klasse, in der die Geschlechtsunterschiede stärker als in 95 % der übrigen Klassen ausgeprägt waren, erhielten die Mädchen Deutschnoten, die um  $d = 0.68$  besser waren als jene der Jungen. Dagegen betrug der Geschlechtsunterschied in der Klasse, in der die Diskrepanzen nur größer als in 5 % der übrigen Klassen waren, nur  $d = 0.15$ . Dies entspricht einer Perzentilbandlänge von mehr als einer halben Standardabweichung der Deutschnote. Geringer fielen die Klassenunterschiede bei der Anspruchsniveausetzung im Fach Deutsch aus, die Länge des Perzentilbands betrug hier  $d = 0.29$ . In immerhin 7 % der Klassen waren die Leistungsaspirationen der Jungen – entgegen des Trends in der Gesamtstichprobe – höher als jene der Mädchen.

Mit einer Perzentilbandlänge von  $d = 0.38$  waren bei der Leistungsbewertung in der ersten

Fremdsprache moderate Klasseneffekte zu verzeichnen. Während in 25 % der Klassen die Leistungsbewertungen der Mädchen um mindestens  $d=0.30$  besser waren als jene der Jungen, waren die Geschlechtsunterschiede in weiteren 24 % der Klassen mit  $d=0.10$  oder weniger nur sehr schwach ausgeprägt oder gar nicht vorhanden. In immerhin 7 % der Klassen erhielten Jungen die besseren Fremdsprachenzensuren. Kleine (aber dennoch signifikante) Klassenunterschiede im Ausmaß der Geschlechterdiskrepanzen zeigten sich hinsichtlich der in der Fremdsprache gerade noch zufriedenstellenden Note (Perzentilbandlänge:  $d=0.21$ ). Entsprechend des nicht signifikanten durchschnittlichen Geschlechtsunterschieds, waren hier höhere Aspirationen der Mädchen in etwa gleich vielen Klassen zu beobachten wie höhere Aspirationen der Jungen (44 % bzw. 56 %). Dennoch zeigten sich in 13 % der Klassen Geschlechtsunterschiede, die mit einer betragsmäßigen Effektgröße von mindestens  $d=0.10$  substantiell waren (höhere Aspirationen der Mädchen: 4 %; höhere Aspirationen der Jungen: 9 %).

#### *Schulunterschiede im Ausmaß von Geschlechtsunterschieden*

Die Analyse von Modell 1 erbrachte für keine der sechs abhängigen Variablen signifikant von Null verschiedene Varianzen  $U_{10}$  der schulspezifischen Regressionsgewichte der Geschlechtsvariable (vgl. Tabelle 4). Allerdings betonen Snijders und Bosker (1999, S. 75), dass die Teststärke des Varianztests deutlich kleiner ist als die Teststärke des Tests der Regressionskoeffizienten, so dass eine theoretisch begründete Hypothese zur Abhängigkeit des entsprechenden Regressionskoeffizienten von Prädiktoren einer höheren Ebene nicht allein auf der Grundlage eines nicht-signifikanten Varianzanteils zurückgewiesen werden darf. Auf der Grundlage dieser Empfehlung prüften wir – wie im folgenden Abschnitt ausgeführt – trotz der nicht-signifikanten Varianzanteile  $U_{10}$  mit Modell 2 die

theoretisch erwarteten Schulformeffekte beim Ausmaß von Geschlechtsunterschieden bei Zensuren und Aspirationen (vertikale Schereneffekte).

#### *Einfluss struktureller Klassen- und Schulmerkmale auf das Ausmaß von Geschlechtsunterschieden*

Die Analyseergebnisse zu den mit Modell 2 geprüften Einflüssen der Geschlechterverteilung und der Größe der Klasse, der Jahrgangsstufe sowie der Schulform auf das Ausmaß von Geschlechtsunterschieden bei Leistungsbewertungen und Leistungsaspirationen sind in Tabelle 5 aufgeführt.

Eine gegenüber Modell 1 verbesserte Modellanpassung erbrachte Modell 2 im Fach Mathematik nur für die Leistungsaspirationen ( $\chi^2(5) = 12.4$ ;  $p < .05$ ), nicht aber für die Leistungsbewertungen ( $\chi^2(5) = 7.5$ ;  $p = .19$ ). Dementsprechend konnte für das klassenspezifische Ausmaß der Geschlechtsunterschiede bei der Mathematikzensur kein Effekt des Mädchenanteils in der Klasse, der Größe der Klasse, der Jahrgangsstufe oder der Schulform nachgewiesen werden. Im Hinblick auf die Stärke der Geschlechterdiskrepanzen bei der Anspruchsniveausetzung im Fach Mathematik fand sich Evidenz für einen temporalen Schereneffekt: Die ohnehin geringeren Leistungsansprüche der Mädchen lagen mit ansteigender Jahrgangsstufe zunehmend unter jenen der Jungen ( $d = -0.04$  pro Jahrgangsstufe; hochgerechnet auf den Unterschied zwischen der 5. und 10. Jahrgangsstufe:  $d = -0.19$ ). Analog zu den Leistungsbewertungen zeigten sich bei den mathematischen Leistungsaspirationen keine Abhängigkeiten von der Geschlechterverteilung, der Klassengröße oder der Schulform.

Für die erhaltene Leistungsbewertung und die Anspruchsniveausetzung im Fach Deutsch erbrachte Modell 2 eine Verbesserung der Modellanpassung gegenüber Modell 1 ( $\chi^2(5) > 13.4$ ;  $p < .05$ ). Bei beiden Maßen war ein temporaler Schereneffekt nachweisbar, wonach Leistungsbewertungen und Leistungs-

aspirationen der Mädchen mit zunehmender Jahrgangsstufe zunehmend besser waren als jene der Jungen. Die Stärke dieses temporalen Schereneffekts lag bei der Deutschzensur bei  $d=0.03$  pro Jahrgangsstufe (Unterschied zwischen 5. und 10. Jahrgangsstufe:  $d=0.16$ ) und beim Anspruchsniveau in Deutsch bei  $d=0.05$  pro Jahrgangsstufe (Unterschied zwischen 5. und 10. Jahrgangsstufe:  $d=0.26$ ). Ebenfalls bei beiden Maßen fand sich Evidenz

für einen vertikalen Schereneffekt: An der Hauptschule und an der Realschule waren die Geschlechterdiskrepanzen bei den Leistungsbewertungen und den Leistungsansprüchen im Fach Deutsch stärker ausgeprägt als am Gymnasium. Mit  $d=0.09$  bis  $d=0.15$  handelte es sich dabei um kleine Schereneffekte. Ebenso wie im Fach Mathematik hatte der Anteil der Mädchen und die Gesamtanzahl der Schüler(innen) in der Klasse auch im

Tabelle 5: Schätzung von Modell 2 zur Prädiktion des klassen- und schulspezifischen Ausmaßes an Geschlechtsunterschieden aus strukturellen Merkmalen von Klassen und Schulen

	Mathematik		Deutsch		Erste Fremdsprache	
	Leistungs- bewertung	Leistungs- aspiration	Leistungs- bewertung	Leistungs- aspiration	Leistungs- bewertung	Leistungs- aspiration
<i>Regressionskoeffizienten</i>						
Achsenabschnitt ( $\gamma_{000}$ )	-0.11	0.10	-0.13	0.08	-0.07	0.08
Weiblich						
Achsenabschnitt ( $\gamma_{100}$ )	-0.16***	-0.24***	0.44***	0.15***	0.21***	0.01
Anteil Weiblich ( $\beta_{11}$ )	0.18	-0.17	0.06	0.11	0.05	-0.02
Klassengröße ( $\beta_{12}$ )	0.01	0.00	0.01	0.00	0.00	0.00
Jahrgangsstufe ( $\beta_{13}$ )	-0.01	-0.04**	0.03**	0.05***	0.04**	0.01
Hauptschule ( $\gamma_{101}$ )	0.01	0.06	0.15**	0.14**	0.18**	0.11*
Realschule ( $\gamma_{102}$ )	0.07	0.02	0.15**	0.09*	0.05	0.00
Jahrgangsstufe ( $\beta_{01}$ )	-0.09***	-0.04***	0.01	-0.05***	-0.09***	-0.04***
Hauptschule ( $\gamma_{001}$ )	-0.52***	0.41***	-0.74***	0.30***	-0.42***	0.34***
Realschule ( $\gamma_{002}$ )	-0.32***	0.10**	-0.47***	-0.01	-0.34***	0.04
Leistungsbewertung ( $\pi_2$ )		0.52***		0.36***		0.54***
<i>Varianzanteile</i>						
Residualvarianz innerhalb Klassen ( $E$ )						
	0.85	0.65	0.74	0.79	0.84	0.65
Residualvarianz zwischen Klassen ( $R_0$ )						
	0.05***	0.03***	0.08***	0.04***	0.06***	0.03***
Residualvarianz zwischen Schulen ( $U_{00}$ )						
	0.01***	0.01***	0.03***	0.01***	0.02***	0.01***
Gewicht Weiblich zwischen Klassen ( $R_1$ )						
	0.04***	0.02***	0.07***	0.03***	0.05***	0.02***
Gewicht Weiblich zwischen Schulen ( $U_{10}$ )						
	0.00	0.00	0.00	0.00	0.00	0.00
<i>Modellanpassung</i>						
Devianz	24937.8	22332.3	23871.3	24155.7	24896.2	22326.1

Anmerkungen: Alle abhängigen Variablen wurden z-standardisiert und so gepolt, dass ein hoher Wert eine gute Leistungsbewertung bzw. eine hohe Leistungsaspiration widerspiegelt. Alle Prädiktoren wurden am Stichprobenmittelwert zentriert.

\*\*\*  $p < .001$  \*\*  $p < .01$  \*  $p < .05$

Fach Deutsch keinen nachweislichen Einfluss auf das Ausmaß der Geschlechtsunterschiede.

Für die erste Fremdsprache konnte mit Modell 2 für die Leistungsbewertung eine signifikant bessere Modellanpassung gegenüber Modell 1 erzielt werden ( $\chi^2(5) = 15.3; p < .01$ ). Für die Anspruchsniveausetzung im Fremdsprachenunterricht gilt dies allerdings nicht ( $\chi^2(5) = 4.9; p = .43$ ). Für die Fremdsprachenzensur, nicht aber für das entsprechende Aspirationsniveau, fand sich Evidenz für einen zum Fach Deutsch analogen temporalen Schereneffekt: Jungen erhielten mit ansteigender Jahrgangsstufe zunehmend schlechtere Fremdsprachennoten als Mädchen ( $d = 0.04$ ; Unterschied zwischen 5. und 10. Jahrgangsstufe:  $d = 0.18$ ). Weiterhin fiel der Geschlechtsunterschied bei Fremdsprachennoten in der Hauptschule um  $d = 0.18$  größer aus als am Gymnasium (Realschule und Gymnasium unterschieden sich nicht signifikant). In abgeschwächter Form fand sich dieser Effekt auch bei der Anspruchsniveausetzung in der ersten Fremdsprache ( $d = 0.11$ ), wobei dieser Effekt vor dem Hintergrund des nicht-signifikanten Devianztests kaum substantiell sein dürfte. Wiederum hatten Geschlechterverteilung und Klassengröße keinen statistisch absicherbaren Einfluss auf das Ausmaß der Geschlechtsunterschiede im Fremdsprachenunterricht.

Die Betrachtung der auch in Modell 2 bei allen sechs abhängigen Variablen noch signifikanten und substantiellen Varianzen  $R_1$  zeigt, dass mit den untersuchten Strukturmerkmalen von Klassen und Schulen allenfalls ein geringer Anteil der Klassenunterschiede im Ausmaß der Geschlechterdiskrepanzen bei Leistungsbewertungen und Leistungsaspirationen aufgeklärt werden konnte. Der Anteil der durch temporale und vertikale Schereneffekte aufgeklärten Varianz  $R_1$  lag bei den Mathematikaspirationen bei 13 %, bei der Deutschnote bei 8 %, bei den Deutschaspirationen bei 21 % und bei der Fremdsprachennote bei 15 %.

## Diskussion

Anliegen der vorliegenden Arbeit war es zunächst, unter Kontrolle von Jahrgangsstufen- und Schulformeffekten eine valide Schätzung durchschnittlicher Geschlechtsunterschiede bei Leistungsbewertungen (Zensuren) in Mathematik, Deutsch und der ersten Fremdsprache und damit in jenen drei Hauptfächern aus dem mathematisch-naturwissenschaftlichen und dem sprachlichen Bereich vorzunehmen, die ab der 5. Jahrgangsstufe unterrichtet werden (Hypothesen  $H_{1a}$  und  $H_{1b}$ ). Erwartungsgemäß und kongruent zur bisherigen Literatur zeigte sich in Mathematik ein kleiner Vorteil von Jungen ( $d = 0.15$ ) und in Deutsch ein mittlerer Vorteil von Mädchen ( $d = 0.42$ ). Die Größe der Effekte lag dabei in dem mit PISA 2003 für die mathematische Kompetenz und die Lesekompetenz gefundenen Bereich (Zimmer et al., 2004). Für die erste Fremdsprache, in den meisten Fällen Englisch, war ebenfalls erwartungsgemäß ein kleiner Vorteil der Mädchen zu beobachten ( $d = 0.17$ ).

Ein ähnliches Muster zeigte sich bei den Leistungsaspirationen (Hypothesen  $H_{1a}$  und  $H_{1b}$ ), allerdings war die Differenz zwischen Jungen und Mädchen bei diesem makroskopischen Parameter der Leistungsmotivation in Mathematik größer und in Deutsch kleiner als bei den Leistungsbewertungen. In der Fremdsprache war kein Geschlechtsunterschied bezüglich der Aspirationen nachweisbar. Offensichtlich zeigen Mädchen fachunabhängig ein niedrigeres Anspruchsniveau als Jungen (vgl. Ziegler & Stöger, 2004). Dieser Befund fügt sich gut in die bestehende Befundlage zu Geschlechtsunterschieden im motivationalen Bereich ein, wonach Mädchen ihre Fähigkeiten niedriger einschätzen als Jungen und eine geringere Erfolgserwartung aufweisen (z. B. Ziegler, Heller & Broome, 1996).

Die zentrale Frage, die wir mit der vorliegenden Arbeit zu beantworten suchten, richtete sich auf die Varianz von Geschlechtsunterschieden zwischen Klassen und Schulen (Hypothesen  $H_{2a}$  und  $H_{2b}$ ). Aufgrund der mul-

tiplen Quellen der geschlechtsspezifischen Sozialisation in der Schulklasse – Lehrkräfte, Mitschüler(innen) und Unterrichtsmedien – nahmen wir an, dass es Unterschiede zwischen Klassen und Schulen im Ausmaß von Geschlechtsunterschieden gibt. Das heißt, dass sowohl Klassen und Schulen existieren, die sich durch stärkere als die geschilderten durchschnittlichen Geschlechtsunterschiede charakterisieren lassen, als auch solche, in denen die durchschnittlichen Geschlechtsunterschiede nicht oder nur abgeschwächt zum Tragen kommen. Für die Klassenebene konnte diese Annahme bestätigt werden, wobei die Varianz zwischen Klassen insbesondere in Bezug auf Leistungsbewertungen substantiell war, aber auch in Bezug auf Leistungsaspirationen gegenüber Zufallsschwankungen statistisch abgesichert werden konnte. Für Leistungsbewertungen und Leistungsaspirationen zeigte sich in allen drei untersuchten Fächern, dass es abweichend vom durchschnittlichen Muster der Geschlechterdiskrepanzen auch Klassen gibt, in denen die Benachteiligung einer Geschlechtergruppe nicht oder nur schwach zum Tragen kommt. In diesen Klassen, die als geschlechtsegalitär bezeichnet werden können, erhalten Jungen und Mädchen vergleichbare Leistungsbewertungen und verfügen somit über vergleichbare Chancen. Gleichzeitig existieren aber auch Klassen, die als besonders geschlechtsdiskriminativ bezeichnet werden können, da hier die Chancen von Mädchen und Jungen sehr deutlich gemäß gängiger Stereotype verteilt und damit besonders ungleich sind. Als entscheidende Ebene für diese Varianz hat sich in unseren Analysen die Klassenebene erwiesen; abgesehen von den erwarteten vertikalen Schereneffekten (Schulformeffekte) fanden wir keine Indizien für eine analoge Varianz zwischen Schulen. Dies ist insofern plausibel, als die Klasse und nicht die gesamte Schule der vorrangige Interaktionsrahmen von Schüler(inne)n ist, in dem geschlechtsspezifische Überzeugungen und Erwartungen sichtbar und wirksam werden.

Teilweise konnten die Klassenunterschiede im Ausmaß von Geschlechterdiskrepanzen durch strukturelle Merkmale von Klassen und Schulen aufgeklärt werden. Bemerkenswert ist dabei aber, dass für alle drei Fächer mit einiger Sicherheit ausgeschlossen werden kann, dass das Ausmaß der Geschlechtsunterschiede im Zusammenhang mit dem Mädchenanteil in der Klasse steht (Hypothese  $H_{3a}$ ). Damit kann beispielsweise ausgeschlossen werden, dass die Geschlechterdiskrepanz in Mathematik dann nivelliert ist, wenn Mädchen in der Klasse in der Überzahl sind. Allerdings ist zu betonen, dass diese Ergebnisse nicht ohne weiteres auf den Fall exklusiver Mädchen- oder Jungenklassen übertragen werden können (vgl. hierzu Kessels, 2002), einerseits weil unsere Stichprobe keine monoedukativen Klassen umfasste, andererseits weil die monoedukative Unterrichtung Erwartungen bei Schüler(inne)n und Lehrkräften auslöst, die bei gemischter Geschlechterverteilung nicht zum Tragen kommen (Ziegler, Broome & Heller, 1998).

Ebenfalls keine Effekte auf das Ausmaß von Geschlechtsunterschieden waren für die Klassengröße nachweisbar. Die Vermutung, dass Lehrkräfte in großen Klassen aufgrund der dortigen höheren kognitiven Belastung (vgl. Schrader et al., 2001) Geschlechtsstereotype bei der Leistungsbewertung besonders stark aktivieren (vgl. Sherman et al., 1998), kann somit zurückgewiesen werden (Hypothese  $H_{3b}$ ).

Dagegen bestätigen unsere Ergebnisse für die beiden sprachlichen Fächer die erwarteten temporalen und vertikalen Schereneffekte (Hypothesen  $H_{3c}$  und  $H_{3d}$ ). Unsere Analysen erbrachten für Leistungsbewertungen in Deutsch und der ersten Fremdsprache Evidenz für einen temporalen Schereneffekt, wonach der Vorsprung der Mädchen in diesen Fächern mit ansteigender Jahrgangsstufe zunimmt. Für die Aspirationen im Fach Deutsch wurde ein analoger Effekt gefunden. Diese Ergebnisse liefern Anhaltspunkte dafür, dass die besonders bei Jungen anzutreffenden Defizite in den

für diese Fächer erforderlichen Kompetenzen (z. B. Kompetenzen im Bereich des Lesens, der Sprachproduktion oder der Textinterpretation) kumulieren und dass eine Förderung möglichst frühzeitig ansetzen sollte. Einen temporalen Schereneffekt im Fach Mathematik konnten wir für das Anspruchsniveau bestätigen, nicht aber für Leistungsbewertungen. Nach unseren Ergebnissen werden die Leistungen von Mädchen in diesem Fach also unabhängig von der Jahrgangsstufe konstant schlechter bewertet als jene von Jungen. Wir konnten ebenfalls Evidenz für einen vertikalen Schereneffekt in den beiden sprachlichen Fächern erbringen. Danach sind die Unterschiede zu Ungunsten von Jungen im Deutschunterricht der Haupt- und der Realschule und im fremdsprachlichen Unterricht der Hauptschule größer als am Gymnasium. Dieser Effekt konvergiert mit dem Ergebnis von PISA 2003, wonach die Lesekompetenz von Jungen auf den untersten Kompetenzstufen besonders drastisch unterhalb der von Mädchen liegt (Zimmer et al., 2004), und markiert hier einen besonderen Förderbedarf.

Im Hinblick auf unsere zentralen Hypothesen  $H_{2a}$  und  $H_{2b}$  ist bedeutsam, dass auch nach Kontrolle der diskutierten Schereneffekte substanziale Klassenunterschiede in der Stärke von Geschlechtsunterschieden bestanden. Diese Varianz zwischen Klassen konnte durch strukturelle Merkmale von Schulklassen und Schulen nur in geringem Maße aufgeklärt werden.

Die von Jahrgangsstufen- und Schulformeffekten unabhängige Varianz im Ausmaß, in dem Geschlechtsunterschiede zum Tragen kommen, kann als weiteres Argument dafür gewertet werden, dass die Ursachen von Geschlechtsunterschieden im Leistungs- und Motivationsbereich zumindest teilweise in einer geschlechtsspezifischen Sozialisation und nicht ausschließlich in anlagebedingten Unterschieden zu suchen sind. Wäre Letzteres der Fall, dürften sich auf Klassenebene keine gegenüber Zufallsschwankungen absicherbaren Unterschiede in der Stärke der Diskrepanzen

finden lassen. Die vorliegenden Daten legen dabei nahe, dass neben anderen Quellen die Schulklasse eine bedeutsame Sozialisationsinstanz darstellt.

Dies lenkt den Blick zunächst auf die Lehrkräfte. Verschiedene Arbeiten belegen, dass diese häufig geschlechtsspezifische Erwartungen und Überzeugungen aufweisen (z. B. Heller, et al., 2000; Rustemeyer, 1999) und Interaktionen mit Schüler(inne)n häufig in geschlechtsspezifischer Weise initiieren (z. B. interagieren sie mit Jungen häufiger als mit Mädchen; Jones & Dindia, 2004). Daneben wurde Evidenz dafür erbracht, dass Jungen eher von einem konkurrenzorientierten und Mädchen eher von einem kooperationsorientierten Klassenklima profitieren (zsf. Beerman et al., 1992). Ein fruchtbares Feld zukünftiger Forschung ist die Überprüfung, ob die Überzeugungen und Verhaltensweisen von Lehrkräften sowie das Klassenklima mit dem Ausmaß von Geschlechtsunterschieden in Klassen kovariieren. Mit dem gleichzeitigen Einbezug von Zensuren und standardisierten Schulleistungstests könnte dabei auch eine Aussage darüber getroffen werden, zu welchem Anteil die von uns identifizierte Varianz zwischen Klassen auf eine Varianz im Ausmaß von Erwartungseffekten bei der Leistungsbewertung durch Lehrkräfte und auf eine Varianz in den tatsächlichen Leistungen zurückgeht. Ergebnisse von standardisierten Schulleistungstests standen für die vorliegende Analyse leider nicht zur Verfügung, was sicherlich eine gewisse Einschränkung darstellt. Ein eindeutiger Rückschluss auf tatsächliche Leistungsunterschiede ist aus den Zensuren aufgrund der genannten Erwartungseffekte nicht möglich. Wohlweislich sprechen wir in der vorliegenden Arbeit deshalb von Leistungsbewertungen und nicht von Leistungen. Gleichwohl sind Zensuren die für Schüler(innen) evidenteste Rückmeldung ihrer Leistung und dürften in ihrer Bedeutung im Motivationsprozess (beispielsweise bei der Entwicklung von Selbstkonzepten; zsf. Möller & Köller, 2004) kaum zu unterschätzen sein.

Neben den Lehrkräfteinflüssen dürfte aber auch der Geschlechtsrollenkonformität von Einstellungen, Überzeugungen und Verhaltensweisen von Mitschüler(inne)n ein bedeutender Einfluss zukommen. In einigen Klassen dürften traditionelle und abwertende Geschlechtsrollenvorstellungen in besonderer Weise aktiviert sein, in anderen Klassen muss dies nicht der Fall sein. Es ist plausibel, dass sich Schüler(innen) unterschiedlicher Klassen darin unterscheiden, wie sehr sie in ihren Wertevorstellungen und Idealselbstbildern den traditionellen Rollenstereotypen verpflichtet sind. Hierbei dürften komplexe Prozesse der Einstellungs- und Werteübernahme eine Rolle spielen, die je nach Klasse in unterschiedlich geschlechtsspezifischen Überzeugungen resultieren. Zur Aufklärung der Varianz von Geschlechtsunterschieden scheint ein genauer Blick auf dieses Interaktionsgeschehen lohnend.

Als pädagogische Ansatzstellen impliziert die von uns identifizizierte Varianz schließlich, dass Lehrkräfte stärker als bisher für mögliche Chancendisparitäten sensibilisiert, über die Ursachen von Geschlechtsunterschieden aufgeklärt und sie zur Herstellung von gleichen Chancen für Mädchen und Jungen befähigt werden sollten. Hierzu erscheinen die Herstellung eines geeigneten Klassenklimas, die Förderung von Selbsteinschätzungen und Motivation bei der Geschlechtergruppe mit den jeweils schlechteren Leistungen sowie die Modellierung von Einstellungen und Verhaltensweisen, die traditionellen Rollenvorstellungen widersprechen, als geeignet.

## Literatur

- Baumert, J., Bos, W. & Lehmann, R. (Hrsg.) (2000). *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie: Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn*. Opladen: Leske + Budrich.
- Beerman, L., Heller, K. A. & Menacher, P. (1992). *Mathe: nichts für Mädchen? Begabung und Geschlecht am Beispiel von Mathematik, Naturwissenschaft und Technik*. Bern: Huber.
- Beyer, S. & Bowden, E. M. (1997). Gender differences in self-perceptions: Convergent evidence from three measures of accuracy and bias. *Personality and Social Psychology Bulletin*, 23, 157–172.
- Bussey, K. & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review*, 106, 676–713.
- Dickhäuser, O. & Plenter, I. (im Druck). „Letztes Halbjahr stand ich zwei“. Zur Akkuratheit selbst berichteter Noten. *Zeitschrift für Pädagogische Psychologie*.
- Dresel, M., Heller, K. A., Schober, B. & Ziegler, A. (2001). Geschlechtsunterschiede im mathematisch-naturwissenschaftlichen Bereich: Motivations- und selbstwertschädliche Einflüsse der Eltern auf Ursachenerklärungen ihrer Kinder in Leistungskontexten. In C. Finkbeiner & G. W. Schnaitmann (Hrsg.), *Lehren und Lernen im Kontext empirischer Forschung und Fachdidaktik* (S. 270–288). Donauwörth: Auer.
- Eagly, A. H., Beall, A. E. & Sternberg, R. J. (2004). *The psychology of gender* (2nd ed.). New York: Guilford Press.
- Eccles, J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. & Midgley, C. (1983). Expectancies, values and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives* (pp. 26–43). San Francisco: Freeman.
- Finsterwald, M. & Ziegler, A. (im Druck). Geschlechtsrollenstereotypen in Schulbüchern: Neue Befunde zu einer alten Frage. In P. Ludwig & H. Ludwig (Hrsg.), *Erwartungen in himmelblau und rosarot – Auslöser lebenslanger Geschlechterdifferenzen im Lernen?* Opladen: Verlag für Sozialwissenschaften.
- Freeman, J. (2004). Cultural Influences on gifted gender achievement. *High Ability Studies*, 15, 7–25.
- Halpern, D. F. (2000). *Sex differences in cognitive abilities* (3rd ed.). Mahwah: Erlbaum.
- Hannover, B. (1991). Zur Unterrepräsentanz von Mädchen in Naturwissenschaften und Technik: Psychologische Prädiktoren der Fach- und Berufswahl. *Zeitschrift für Pädagogische Psychologie*, 5, 169–186.
- Hannover, B. (2004). Gender revisited: Konsequenzen aus PISA für die Geschlechterforschung. *Zeitschrift für Erziehungswissenschaften*, 7, 81–99.
- Hannover, B. & Kessels, U. (2002). Monoedukativer Anfangsunterricht in Physik in der Gesamtschule. Auswirkungen auf Motivation, Selbstkonzept und Einteilung in Grund- und Fortgeschrittenenurse. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 34, 201–215.
- Heckhausen, H. (1989). *Motivation und Handeln* (2. Aufl.). Berlin: Springer.
- Heller, K. A., Finsterwald, M. & Ziegler, A. (2000). Implicit theories of German Mathematics and Physics teachers on gender specific giftedness and motivation. *Psychologische Beiträge*, 43, 172–189.
- Hill, J. & Lynch, M. (1983). The intensification of gender-related role expectations during early adolescence. In J. Brooks-Gunn & A. Petersen (Eds.), *Girls at puberty: Biological and psychological perspectives* (pp. 201–228). New York: Plenum.
- Hoffmann, L., Häußler, P. & Peters-Haft, S. (1997). *An den Interessen von Mädchen und Jungen orientierter Physikunterricht. Ergebnisse eines BLK-Modellversuchs*. Kiel: Institut für die Pädagogik der Naturwissenschaften.
- Hosenfeld, I., Köller, O. & Baumert, J. (1999). Why sex differences in Mathematics achievement disappear in German secondary schools: A reanalysis of the German TIMMS-data. *Studies in Educational Evaluation*, 25, 143–161.

- Hosenfeld, I., Helmke, A., Ridder, A. & Schrader, F.-W. (2001). Eine mehrbenenanalytische Betrachtung von Schul- und Klasseneffekten. *Empirische Pädagogik, 15*, 513–534.
- Jones, S. M. & Dindia, K. (2004). A meta-analytic perspective on sex equity in the classroom. *Review of Educational Research, 74*, 443–471.
- Kerr, B. A. (1994). *Smart girls two: A new psychology of girls, women, and giftedness*. Dayton: Ohio Psychology Press.
- Kessels, U. (2002). *Undoing Gender in der Schule. Eine empirische Studie über Koedukation und Geschlechtsidentität*. Weinheim: Beltz.
- Lewin, K., Dembo, T., Festinger, L. & Sears, P. S. (1944). Level of aspiration. In J. McHunt (Ed.), *Personality and the behavior disorders* (Vol. 1, pp. 333–378). New York: Ronald Press.
- Middleton, J. & Spanias, P. (1999). Motivation for achievement in mathematics: Findings, generalizations, and criticisms of the research. *Journal for Research in Mathematics Education, 30*, 65–88.
- Möller, J. & Köller, O. (2004). Die Genese akademischer Selbstkonzepte: Effekte dimensionaler und sozialer Vergleiche. *Psychologische Rundschau, 55*, 19–27.
- Pintrich, P. R. & Schunk, D. H. (2002). *Motivation in education: Theory, research and application* (2nd ed.). Englewood Cliffs, NJ: Merrill Prentice Hall.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F. & Congdon, R. T. (2001). *HLM 5* [Computer Software, Version 5.04]. Chicago: Scientific Software International.
- Rustemeyer, R. & Jubel, A. (1996). Geschlechtsspezifische Unterschiede im Unterrichtsfach Mathematik hinsichtlich der Fähigkeitseinschätzung, Leistungserwartung, Attribution sowie im Lernaufwand und im Interesse. *Zeitschrift für Pädagogische Psychologie, 10*, 13–25.
- Rustemeyer, R. (1999). Geschlechtstypische Erwartungen zukünftiger Lehrkräfte bezüglich des Unterrichtsfaches Mathematik und korrespondierende (Selbst-)Einschätzungen von Schülerinnen und Schülern. *Psychologie in Erziehung und Unterricht, 46*, 187–200.
- Schrader, F.-W., Helmke, A., Hosenfeld, I. & Ridder, A. (2001). Klassengröße und Mathematikleistung. *Empirische Pädagogik, 15*, 601–625.
- Sherman, J. W., Lee, A. Y., Bessenoff, G. R. & Frost, L. A. (1998). Stereotype efficiency reconsidered: Encoding flexibility under cognitive load. *Journal of Personality and Social Psychology, 75*, 589–606.
- Snijders, T. A. B. & Bosker, R. J. (1999). *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. London: Sage.
- 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. 249–269). Opladen: Leske + Budrich.
- Stanat, P. & Kunter, M. (2003). Kompetenzerwerb, Bildungsbeteiligung und Schullaufbahn von Mädchen und Jungen im Ländervergleich. 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. 211–242). Opladen: Leske + Budrich.
- Tiedemann, J. & Faber, G. (1994). Mädchen und Grundschulmathematik: Ergebnisse einer vierjährigen Längsschnittuntersuchung zu ausgewählten geschlechtsbezogenen Unterschieden in der Leistungsentwicklung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 26*, 101–111.
- Trautner, H. M. (1994). Geschlechtsspezifische Erziehung und Sozialisation. In K. A. Schneewind (Hrsg.), *Psychologie der Erziehung und Sozialisation* (S. 167–195). Göttingen: Hogrefe.
- Trautner, H. M., Helbing, N., Sahn, W. B. & Lohaus, A. (1988). Unkenntnis – Rigidität – Flexibilität: Ein Entwicklungsmodell der Geschlechtsrollen-Stereotypisierung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 19*, 105–120.
- Wang, M. C., Haertel, G. D. & Walberg, H. J. (1990). What influences learning? A content analysis of review literature. *Journal of Educational Research, 84*, 30–43.
- Ziegler, A., Broome, P. & Heller, K. A. (1998). Pygmalion im Mädchenkopf. Erwartungs- und Erfahrungseffekte koedukativen vs. geschlechtshomogenen Physikanfangsunterrichts. *Psychologie in Erziehung und Unterricht, 45*, 2–18.
- Ziegler, A., Dresel, M. & Schober, B. (2000). Prädiktoren des Selbstvertrauens von Mädchen und Jungen vor dem erstmaligen Chemieunterricht am Gymnasium. *Psychologie in Erziehung und Unterricht, 47*, 66–75.
- Ziegler, A., Dresel, M., Schober, B. & Stöger, H. (in Vorb.). *Die Ulmer Motivationstestbatterie (UMTB)*.
- Ziegler, A., Heller, K. A. & Broome, P. (1996). Motivational preconditions of gifted and highly gifted girls in physics. *High Ability Studies, 7*, 129–143.
- Ziegler, A. & Stöger, H. (2004). Differential effects of motivational orientation on self-confidence and helplessness among high achievers and underachievers. *Gifted and Talented International, 19*, 61–68.
- 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.

Dr. Markus Dresel  
 Abt. Pädagogische Psychologie  
 Universität Ulm  
 D-89069 Ulm  
 Tel.: (07 31) 50-3 11 32  
 Fax: (07 31) 50-3 11 37  
 E-Mail: markus.dresel@uni-ulm.de