

Running Head: Dreiebenenanalyse von Geschlechtsunterschieden

Geschlechtsunterschiede bei Leistungsbewertungen und Leistungsaspirationen: Eine
Dreiebenenanalyse zu Klassen- und Schulunterschieden im Ausmaß von
Geschlechterdifferenzen

Markus Dresel, Heidrun Stöger & Albert Ziegler

Fassung vom 13. März 2005

Autorenhinweis

Dr. Markus Dresel, Dr. Heidrun Stöger, Prof. Dr. Albert Ziegler, Universität Ulm
Korrespondenzadresse: Dr. Markus Dresel, Abt. Pädagogische Psychologie, Universität Ulm,
D-89069 Ulm, E-Mail: markus.dresel@sem-paedagogik.uni-ulm.de

Abstract

An einer Stichprobe von 9.207 Schüler(inne)n aus 498 Klassen der 5. bis 10. Jahrgangsstufe an 54 koedukativen Schulen wurden Geschlechtsunterschiede bei Leistungsbewertungen (Zensuren) und Leistungsaspirationen unter Einbeziehung von Individual-, Klassen- und Schulebene analysiert. Nachgewiesen wurden Vorteile von Jungen im Mathematik- und Vorteile von Mädchen im Deutsch- und Fremdsprachenunterricht. In allen drei Schulfächern konnten im Ausmaß der Geschlechtsunterschiede bei Leistungsbewertungen und – aspirationen substanzielle Unterschiede zwischen den untersuchten Klassenräumen identifiziert werden: Neben Klassen, in denen die Diskrepanzen zwischen den Geschlechtern stark ausgeprägt waren, fanden sich auch Klassen in denen diese nur abgeschwächt oder gar nicht zum Tragen kamen. Daneben wurde für den Deutsch- und den Fremdsprachenunterricht Evidenz erbracht sowohl für einen temporalen Schereneffekt, wonach der Vorteil der Mädchen mit zunehmender Jahrgangsstufe an Substanz gewinnen, als auch für einen vertikalen Schereneffekt, wonach diese Geschlechtsunterschiede in Haupt- und Realschule stärker ausgeprägt sind als am Gymnasium.

Schlüsselwörter: Geschlechtsunterschiede, Schulleistung, Motivation, Klassenunterschiede

Geschlechtsunterschiede im schulischen Leistungsbereich sind ein seit langer Zeit diskutiertes Problemfeld. Breit beforscht ist insbesondere der mathematisch-naturwissenschaftliche Bereich, wobei sich Schulleistungsunterschiede zu Ungunsten der Mädchen ab der Sekundarstufe zeigten, und zwar weitgehend unabhängig davon, ob Leistungsbewertungen durch Lehrkräfte (Zensuren) oder kriteriale Schulleistungstests Verwendung fanden wurden (z.B. Broome, 1998; Middleton & Spanias, 1999; Stanat & Kunter, 2001). Basierend auf der Beobachtung, dass sich schlechtere mathematische Schulleistungen von Mädchen in der Grundschule noch nicht nachweisen lassen (Tiedemann & Faber, 1994; Marshall & Smith, 1987), wird ein Schereneffekt vermutet, wonach die Geschlechtsunterschiede zum Nachteil der Mädchen im Laufe der Schulzeit an Substanz gewinnen. Daneben ist in mathematisch-naturwissenschaftlichen Fächern auch die Motivation von Mädchen deutlich ungünstiger ausgeprägt: Im Mittel haben sie ein niedrigeres Fachinteresse, geringe Erfolgserwartungen, ein niedrigeres Fähigkeitsselbstkonzept und erklären sich das Zustandekommen von Erfolgen und Misserfolgen motivationsabträglicher und selbstwertgefährdender als Jungen (z.B. Beyer & Bowden, 1997; Hannover, 1991; Hoffmann, Häußler & Peters-Haft, 1997; Rustemeyer & Jubel, 1996; Stipek & Gralinski, 1991; Ziegler, Dresel & Schober, 2000). Entsprechend der durchschnittlich schlechteren Leistungen und ungünstigeren Motivation haben Mädchen in mathematisch-naturwissenschaftlichen Domänen niedrigere Leistungsaspirationen, geben sich also mit schlechteren Leistungsergebnissen zufriedener als Jungen (zsf. Beerman, Heller & Menacher, 1992).

Die internationale Schulleistungstudien PISA 2000 und PISA 2003 bestätigten für den mathematischen Bereich einen Vorteil der Jungen (Stanat & Kunter, 2001; 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 deutliche Vorteile der Mädchen bei der Lesekompetenz (Zimmer et al., 2004). Die Effektstärke des Unterschieds lag im OECD-Durchschnitt bei $d=-0.35$, in Deutschland bei $d=-0.39$. Im Vergleich dazu fiel der Unterschied in der mathematischen Basiskompetenz mit $d=0.11$ im OECD-Durchschnitt und $d=0.09$ in Deutschland deutlich geringer aus. Sowohl bei der Lesekompetenz als auch der mathematischen Kompetenz zeigte PISA, dass das Ausmaß der Geschlechtsunterschiede in den 29 OECD-Staaten sehr unterschiedlich ist. Nach Zimmer et al. (2004) ist der Vorteil der Mädchen bei der Lesekompetenz zwar in allen Staaten signifikant, schwankt aber zwischen $d=-0.61$ (Island) und $d=-0.21$ (Japan). Signifikante Vorteile der Jungen im mathematischen Kompetenzbereich berichten die Autoren nur für 21

Staaten; hier schwankt der Geschlechtsunterschied zwischen $d=0.26$ (Korea) und $d=-0.17$ (Island).

Die Geschlechtsunterschiede im sprachlichen Leistungsbereich zu Ungunsten der Jungen blieben lange Zeit unbeachtet – entsprechend groß ist noch das diesbezügliche Forschungsdefizit. Beispielsweise liegen keine systematischen Erkenntnisse zur Größe der Leistungsunterschiede in verschiedenen Jahrgangsstufen vor – PISA untersuchte nur Schüler(innen) im Alter von 15 Jahren –, so dass bislang offen ist, wann die Vorteile der Mädchen im sprachlichen Bereich zutage treten und ob diese Vorteile im zeitlichen Verlauf konstant sind oder im Sinne eines Schereneffekts an Substanz gewinnen. Zudem fehlen repräsentative Aussagen zum Ausmaß von Geschlechtsunterschieden bei Zensuren in sprachlichen Fächern, also bei summarischen Indikatoren der Schulleistung, für die neben der Lesekompetenz auch andere Kompetenzen von Bedeutung sind (z.B. Kompetenzen im Bereich der Sprachproduktion oder der Textinterpretation). Eine Differenzierung nach Schulformen und nach einzelnen Schulfächern, insbesondere nach muttersprachlichem und fremdsprachlichem Unterricht, dürfte zusätzlichen Aufschluss über die Charakteristik von Geschlechterdifferenzen im sprachlichen Bereich liefern. Nicht zuletzt mangelt es an Befunden zur Motivation von Schüler(inne)n in sprachlichen Domänen.

Die meisten theoretischen Erklärungen der Geschlechtsunterschiede heben auf eine geschlechtsspezifische Sozialisation ab (z.B. Halpern, 2000; Trautner, 1994). In vielen Modellen ist die Annahme enthalten, 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, deren Erwartungen und Überzeugungen erheblichen Einfluss auf die Herausbildung geschlechtsspezifischer Motivationen haben (z.B. Dresel, Heller, Schober & Ziegler, 2001), der Klassenraum eine bedeutsame Rolle spielt. Für den mathematisch-naturwissenschaftlichen Bereich ist beispielsweise gut belegt, dass Lehrkräfte häufig geschlechtsspezifische Erwartungen und Überzeugungen aufweisen, wonach hier Jungen weniger Schwierigkeiten haben, effektiver lernen und generell begabter sind als Mädchen (z.B. Heller, Finsterwald, & Ziegler, 2000; Rustemeyer, 1999). Diese Muster finden ihre Entsprechung in geschlechtsspezifischen Interaktionsstilen in Unterricht und Schule (zsf. Jones & Dindia, 2004). Daneben können Überzeugungen und Wertesysteme von Mitschüler(inne)n vorhandene geschlechtsbezogene Selbstbilder noch verstärken und

stabilisieren. Schließlich finden sich auch in Schulbüchern nach wie vor traditionelle Geschlechtsrollenstereotype (Finsterwald & Ziegler, 2005).

Entsprechend dieser multiplen Bedingungskonstellation ist nicht davon auszugehen, dass der Einfluss dieser Sozialisationsinstanz über alle Klassen und Schulen hinweg konstant ist. Vielmehr kann angenommen werden, dass sich Klassenräume dahingehend unterscheiden, wie stark geschlechtsspezifische Einstellungen, Überzeugungen und Interaktionen vorzufinden und in Lernen und Leistung wirksam werden. Dementsprechend vermuten wir, dass eine substantielle *Variabilität im Ausmaß der Geschlechtsunterschiede zwischen Klassen und Schulen* besteht. Diese Vermutung entspricht auf niedriger Ebene den relativ großen Schwankungen im Ausmaß der Geschlechtsunterschiede, die mit PISA zwischen den beteiligten Staaten gefunden und oben beschrieben wurden. Im Interpretationsrahmen sozialisationstheoretischer Erklärungsansätze bedeuten diese Schwankungen, dass die Sozialisationserfahrungen von Jungen und Mädchen je nach kulturellem Kontext in unterschiedlich starkem Ausmaß durch differierende Rollenerwartungen und Stereotype charakterisiert sind. In analoger Weise mag es „geschlechtsprogressive“ Klassen geben, in denen die bekannten Geschlechtsunterschiede in Leistung und Motivation nicht oder nur abgeschwächt zum Tragen kommen und die im Kontrast zu „geschlechtskonservativen“ Klassen stehen, in denen diese Differenzen zwischen den Geschlechtern besonders stark ausgeprägt sind. Diese Variabilität zwischen Klassen (und möglicherweise ganzen Schulen) im Ausmaß von Leistungs- und Motivationsunterschieden wurde bislang allerdings nicht untersucht.

Um Unterschiede zwischen Klassen in Schule zu untersuchen, ist eine mehrebenenanalytische Betrachtung erforderlich, die eine separate Schätzung der Varianz auf Individual-, Klassen- und Schulebene erlaubt. Insbesondere ist es nur damit möglich, die interessierende Variabilität zwischen Klassen und Schulen im Hinblick auf die Größe des Geschlechtseffekts zu schätzen. Daneben bietet eine Dreiebenenmodellierung unter Einbezug von Individual-, Klassen- und Schulebene auch die Chance, simultan zu den Einflüssen der individuellen Geschlechtsvariable Einflüsse unterschiedlicher Geschlechterverteilungen in Klassen zu analysieren. In diesem Zusammenhang wäre es beispielsweise denkbar, dass bei unausgewogenem Geschlechterverhältnis gegenüber einer Minderheit mit schlechteren Leistungen ein Mildeeffekt bei der Leistungsbewertung durch die Lehrkräfte wirksam wird. Danach sollte etwa in Klassen mit geringem Mädchenanteil der Geschlechtsunterschied im Fach Mathematik zu Ungunsten der Mädchen abgeschwächt sein. Andererseits könnte in solchen Klassen den Leistungen der überzähligen Jungen eine normative Funktion zukommen

und als Referenz für die Bewertung der (durchschnittlich schlechteren) Leistungen der Mädchen fungieren, was zu einer Vergrößerung der Geschlechtsunterschiede führen sollte. Daneben ist es aber auch denkbar, dass die Geschlechterverteilung einen für alle Schüler(innen) einer Klasse wirksamen Effekt bei der Bewertung von Leistungen hat. So könnte eine Klasse mit hohem Mädchenanteil von Lehrkräften als „gute Klasse“ in sprachlichen Fächern und als „schlechte Klasse“ in mathematisch-naturwissenschaftlichen Fächern angesehen werden, was aufgrund von Erwartungseffekten zu insgesamt besseren Leistungsbewertungen für Jungen und Mädchen in Deutsch und fremdsprachlichen Fächern sowie zu insgesamt schlechteren Bewertungen bei Mädchen und Jungen in Mathematik und naturwissenschaftlichen Fächern führen könnte.

Fragestellungen

Die zentrale Fragestellung der vorliegenden Arbeit richtet sich darauf, ob sich *Klassen und Schulen dahingehend unterscheiden, wie stark das Ausmaß von Geschlechtsunterschieden bei der Schulleistung* ist. Um differenzierte Aussagen zu Benachteiligungen von Mädchen und Jungen ableiten zu können fokussieren wir mit Mathematik, Deutsch und der ersten Fremdsprache *Fächer sowohl aus dem mathematisch-naturwissenschaftlichen als auch dem sprachlichen Bereich*. Als Indikator der Schulleistung beziehen wir Leistungsbewertungen durch Lehrkräfte (*Zensuren*) ein. Um daneben auch Anhaltspunkte zu Klassen- und Schuleffekten in Bezug auf Geschlechtsunterschiede im motivationalen Bereich zu erhalten, betrachten wir als klassischen Indikator für die Motivation der Schüler(innen) deren *Leistungsaspirationen* (Atkinson & Litwin, 1960; Lewin, Dembo, Festinger & Sears, 1944; Übersicht bei Heckhausen, 1989).

Da Geschlechtsunterschiede durch Jahrgangsstufen und Schulformeffekte überformt oder verdeckt sein können (vgl. Zimmer et al., 2004, S. 212f.), ist es zur validen Schätzung von Geschlechterdifferenzen und von Zwischenklassen- und Zwischenschulvarianzen im Ausmaß von Geschlechtsunterschieden erforderlich, einen breiten Jahrgangsstufenbereich und alle wesentlichen Schulformen einzubeziehen und die entsprechenden Effekte zu kontrollieren. Dies ermöglicht auch die Beantwortung der bislang nicht systematisch untersuchten Frage nach möglichen *temporalen Schereneffekten*, wonach die Geschlechtsunterschiede mit ansteigender Jahrgangsstufe an Substanz gewinnen. Schließlich soll auch analysiert werden, ob die Geschlechterdifferenzen schulformspezifisch ausfallen und damit möglicherweise *vertikale Schereneffekte* vorliegen.

Im Falle substanzieller Klassen- oder Schuleffekte stellt sich die Frage nach den *Prädiktoren des klassen- bzw. schulspezifischen Ausmaß an Geschlechtsunterschieden*. In der vorliegenden Arbeit betrachten wir dazu strukturelle Merkmale von Klassen und Schulen.

Methode

Es wurde ein Teil eines größeren Datensatzes analysiert, der zur Normierung der sich in Entwicklung befindlichen Ulmer Motivationstestbatterie (*UMTB*; vgl. Dresel, Schober, Stöger & Ziegler, 2005) gewonnen wurde und Daten von Schüler(inne)n aus Hauptschul-, Realschul- und Gymnasialklassen der 5. bis 10. Jahrgangsstufen enthält. Als abhängige Variablen wurden Leistungsbewertungen (Zensuren) und Leistungsaspirationen in drei Hauptfächern (Mathematik, Deutsch, erste Fremdsprache) in die Analyse einbezogen. Als unabhängige Variablen fanden biografische Angaben sowie strukturelle Merkmale der Schulklassen und Schulen Verwendung (Geschlecht, Jahrgangsstufe, Geschlechterverhältnis in der Klasse, Klassengröße, Schulform). Geschlechtsunterschiede in den drei Schulfächern wurden hierarchisch linear auf drei Ebenen modelliert (Individual-, Klassen- und Schulebene).

Stichprobe

Die Stichprobe umfasste 4.479 Jungen und 4.728 Mädchen aus 498 Klassen der 5. bis 10. Jahrgangsstufe (JGS) an 54 koedukativen Schulen (5. JGS: 19.5%. 6. JGS: 19.4%. 7. JGS: 18.9%. 8. JGS: 16.8%. 9. JGS: 16.5%. 10. JGS: 8.9%). Die Schüler(innen) nahmen freiwillig und mit Einverständnis ihrer Eltern an der Befragung teil. Deren Durchschnittsalter betrug zum Zeitpunkt der Befragung 13.6 Jahre ($SD=1.76$). Alle teilnehmenden Schulen sind in Baden-Württemberg angesiedelt, wobei 162 Klassen an Hauptschulen, 141 Klassen an Realschulen und 195 Klassen an Gymnasien geführt wurden. Die Größe der Klassen differierte zwischen 10 und 32 Schüler(inne)n ($M=18.5$; $SD=5.3$), die Anzahl der Klassen pro Schule lag in unserer Stichprobe zwischen 1 und 29 ($M=9.2$; $SD=6.6$).¹ Der Mädchenanteil in den Klassen variierte zwischen 8% und 94% ($M=52\%$; $SD=15\%$) und unterschied sich – entsprechend der Verteilung in der Grundgesamtheit – systematisch zwischen den Schulformen: Der Mädchenanteil lag in Hauptschulklassen durchschnittlich bei 48.3%, in Realschulklassen bei 50.4% und in Gymnasialklassen bei 55.1% (keine Unterschiede zwischen verschiedenen Jahrgangsstufen). Die überwiegende Zahl der Schüler(innen) hatte

Englisch (96.0%), lediglich 2.7% hatte Latein und 1.3% hatte Französisch als erste Fremdsprache.

Erfasste Maße

Leistungsbewertungen. Die Zensuren in den Fächern Mathematik und Deutsch wurde mit je einem Item erfasst („Welche Mathematik-/Deutsch-Note hattest du im letzten Zeugnis?“). Zur Erfassung der Leistungsbewertung in der ersten Fremdsprache wurde zunächst erfragt, welche dies ist („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 fünften Jahrgangsstufe wurden um die Angabe der Zensuren in der letzten Klassenarbeit gebeten, da das vorangegangene Zeugnis bei diesen Schüler(inne)n möglicherweise das Jahreszeugnis der vierten Jahrgangsstufe ist und damit in den wenigsten Fällen eine Fremdsprachenote aufweist.

Leistungsaspirationen. Als Maß für die Leistungsaspirationen in den drei Hauptfächern kamen ebenfalls Notenskalen zum Einsatz. Erfasst wurde die schlechteste, noch zufriedenstellende Zensur („Mit welcher Note in Mathematik-/Deutsch-Klassenarbeiten bist du gerade noch zufrieden?“ und „Mit welcher Note in Klassenarbeiten bist du in dieser Fremdsprache gerade noch zufrieden?“).

Analysemethodik

Es wurden mehrere hierarchisch lineare Regressionsmodelle mit drei Analyseebenen (Individualebene, Klassenebene, Schulebene) spezifiziert und mit *HLM 5* (Raudenbush, Bryk, Cheong & Congdon, 2001) getestet. Zunächst wurde das leere Modell (Nullmodell) formuliert, um Zwischenklassen- und Zwischenschulvarianzen der abhängigen Variablen schätzen zu können. Hier wird die abhängige Variable Y_{ijk} des/der i -ten Schüler(in) in der j -ten Klasse in der k -ten Schule durch Achsenabschnittsparameter und Residuen auf allen drei Ebenen ausgedrückt:

Nullmodell:

$$\text{Individualebene: } Y_{ijk} = \boldsymbol{\mu}_{0,jk} + e_{ijk} \quad (1)$$

$$\text{Klassenebene:} \quad \mathbf{p}_{0jk} = \mathbf{b}_{00k} + r_{0jk} \quad (2)$$

$$\text{Schulebene:} \quad \mathbf{b}_{00k} = \mathbf{g}_{000} + u_{00k} \quad (3)$$

Gleichung 1 ist eine übliche multiple Regressionsgleichung, mit dem Unterschied, dass der Achsenabschnittsparameter \mathbf{p}_{0jk} getrennt für jede Klasse j geschätzt wird und als Klassenmittelwert der abhängigen Variablen interpretiert werden kann (individuelles Residuum innerhalb der Klasse: e_{ijk}). Dazu wird für diesen eine weitere Regressionsgleichung formuliert (Gleichung 2): Der Parameter \mathbf{p}_{0jk} wird in Abhängigkeit eines weiteren Achsenabschnittsparameter \mathbf{b}_{00k} , der getrennt für jede Schule k geschätzt wird und als Schulmittelwert interpretiert werden kann, sowie eines klassenspezifischen Residualanteils r_{0jk} formuliert. Schließlich wird für den schulspezifischen Parameter \mathbf{b}_{00k} eine dritte Gleichung spezifiziert, in der dieser als Summe eines übergreifenden Achsenabschnittsparameters \mathbf{g}_{000} (interpretierbar als Stichprobenmittelwert) sowie eines schulspezifischen Residuums u_{00k} aufgefasst wird. Durch diese Modellspezifikation wird es möglich, die Gesamtvarianz der abhängigen Variablen in Bezug auf die drei Ebenen zu dekomponieren, wobei gilt:

$$\text{Var}(Y_{ijk}) = \text{Var}(e_{ijk}) + \text{Var}(r_{0jk}) + \text{Var}(u_{00k}) \quad (4)$$

$E = \text{Var}(e_{ijk})$ entspricht dabei der Varianz innerhalb von Klassen, $R_0 = \text{Var}(r_{0jk})$ der Varianz zwischen Klassen aber innerhalb von Schulen und $U_{00} = \text{Var}(u_{00k})$ der Varianz zwischen Schulen.

Um eine von Jahrgangsstufen- und Schulformeffekten unabhängige Schätzung der Geschlechterdiskrepanzen zu erhalten, wurde im nächsten Schritt ein Modell 1 formuliert, das in den folgenden Gleichungen repräsentiert ist:

Modell 1:

$$\text{Individualebene:} \quad Y_{ijk} = \mathbf{p}_{0jk} + \mathbf{p}_{1jk} \cdot \text{Weiblich} + e_{ijk} \quad (5)$$

$$\text{Klassenebene:} \quad \mathbf{p}_{0jk} = \mathbf{b}_{00k} + \mathbf{b}_{01k} \cdot \text{AnteilWeiblich} + \mathbf{b}_{02k} \cdot \text{JGS} + r_{0jk} \quad (6)$$

$$\mathbf{p}_{1jk} = \mathbf{b}_{10k} + r_{1jk} \quad (7)$$

$$\text{Schulebene:} \quad \mathbf{b}_{00k} = \mathbf{g}_{000} + \mathbf{g}_{001} \cdot \text{HS} + \mathbf{g}_{002} \cdot \text{RS} + u_{00k} \quad (8)$$

$$\mathbf{b}_{01k} = \mathbf{g}_{010} + u_{01k} \quad (9)$$

$$\mathbf{b}_{02k} = \mathbf{g}_{020} \quad (10)$$

$$\mathbf{b}_{10k} = \mathbf{g}_{100} + u_{10k} \quad (11)$$

In dieses Modell wurden zunächst das Geschlecht auf Individualebene (repräsentiert als Dummyvariable *Weiblich* mit männlich als Referenzkategorie) und der Mädchenanteil auf Klassenebene (*AnteilWeiblich*; variiert im Bereich 0.00-1.00) als Prädiktoren eingesetzt (Gleichungen 5 und 6). Zur Auspartialisierung von Jahrgangsstufen- und Schulformeffekten wurden daneben die Jahrgangsstufe (*JGS*) auf Klassenebene und die Schulform (repräsentiert als zwei Dummyvariablen *HS* und *RS* für Hauptschule und Realschule mit Gymnasium als Referenzkategorie) auf Schulebene in Modell 1 eingesetzt (Gleichungen 6 und 8). Der Effekt der Jahrgangsstufe wurde auf Schulebene fixiert ($u_{02k}=0$; Gleichung 10).

Das Regressionsgewicht der Geschlechtsvariable wird in Modell 1 noch als unabhängig von Klassen- und Schulmerkmalen gesehen: In den entsprechenden Gleichungen 7 und 11 werden lediglich die Achsenabschnittsparameter b_{10k} und \mathbf{g}_{100} sowie die Residualanteile r_{1jk} und u_{10k} spezifiziert. Ebenso wird der Einfluss des Mädchenanteils in der Klasse als unabhängig von Schulmerkmalen betrachtet (Spezifikation des Achsenabschnittsparameters \mathbf{g}_{010} und des schulspezifischen Residuums u_{01k} in Gleichung 9). Damit schätzt der Koeffizienten \mathbf{g}_{100} die mittlere Differenz der abhängigen Variablen zwischen Mädchen und Jungen und der Koeffizienten \mathbf{g}_{010} die mittlere Differenz zwischen reinen Mädchenklassen und reinen Jungenklassen (als theoretischer Wert an dem die wahre Variabilität der Geschlechterverteilungen zu relativieren ist – die analysierte Stichprobe enthält keine monoedukativen Klassen). Neben diesen Parametern der mittleren Geschlechtseffekte stellen die Varianzen der Residualanteile $R_1=Var(r_{1jk})$, $U_{10}=Var(u_{10k})$ und $U_{01}=Var(u_{01k})$ bei der Analyse von Modell 1 die eigentlich interessierenden Größen dar. Sie schätzen die Variabilität der Parameter \mathbf{g}_{100} und \mathbf{g}_{010} und geben damit an, wie stark der Einfluss des Geschlechts zwischen Klassen und Schulen variiert.

Daneben erlaubt Modell 1 Aussagen zum Zusammenhang von Jahrgangsstufe (\mathbf{g}_{020}) und Schulform (\mathbf{g}_{001} und \mathbf{g}_{002}) mit den abhängigen Variablen (Gleichungen 6, 8 und 10). Der Parameter \mathbf{g}_{000} gibt die mittlere und um die Einflüsse der genannten Prädiktoren korrigierte Ausprägung der abhängigen Variablen für Schüler(innen) von Gymnasien an (Referenzkategorie der Dummyvariablen). Die Residualvarianzen $R_0=Var(r_{0jk})$ und $U_{00}=Var(u_{00k})$ geben schließlich an, wie stark die abhängigen Variablen noch nach Berücksichtigung aller Prädiktoren zwischen Klassen bzw. Schulen variieren.

Zur Erklärung von etwaigen Klassen- und Schulunterschieden im Ausmaß der Geschlechtsunterschiede wurde Modell 1 um Prädiktoren der Regressionsgewichte \mathbf{p}_{1jk} , \mathbf{b}_{10k} und \mathbf{b}_{01k} zu einem sog. *Slopes-as-Outcomes-Model* erweitert. Die Gleichungen dieses Modells 2 lauten:

Modell 2:

$$\text{Individualebene: } Y_{ijk} = \mathbf{p}_{0jk} + \mathbf{p}_{1jk} \cdot \text{Weiblich} + e_{ijk} \quad (5)$$

$$\text{Klassenebene: } \mathbf{p}_{0jk} = \mathbf{b}_{00k} + \mathbf{b}_{01k} \cdot \text{AnteilWeiblich} + \mathbf{b}_{02k} \cdot \text{JGS} + r_{0jk} \quad (6)$$

$$\mathbf{p}_{1jk} = \mathbf{b}_{10k} + \mathbf{b}_{11k} \cdot \text{AnteilWeiblich} + \mathbf{b}_{12k} \cdot \text{JGS} + \mathbf{b}_{13k} \cdot \text{KGröße} + r_{1jk} \quad (7')$$

$$\text{Schulebene: } \mathbf{b}_{00k} = \mathbf{g}_{000} + \mathbf{g}_{001} \cdot \text{HS} + \mathbf{g}_{002} \cdot \text{RS} + u_{00k} \quad (8)$$

$$\mathbf{b}_{01k} = \mathbf{g}_{010} + \mathbf{g}_{011} \cdot \text{HS} + \mathbf{g}_{012} \cdot \text{RS} + u_{01k} \quad (9')$$

$$\mathbf{b}_{02k} = \mathbf{g}_{020} \quad (10)$$

$$\mathbf{b}_{10k} = \mathbf{g}_{100} + \mathbf{g}_{101} \cdot \text{HS} + \mathbf{g}_{102} \cdot \text{RS} + u_{10k} \quad (11')$$

$$\mathbf{b}_{11k} = \mathbf{g}_{110} \quad (12)$$

$$\mathbf{b}_{12k} = \mathbf{g}_{120} \quad (13)$$

$$\mathbf{b}_{13k} = \mathbf{g}_{130} \quad (14)$$

Als Prädiktoren des klassen- und schulspezifischen Einflusses der individuellen Geschlechtsvariable wurden die Klassenmerkmale „Mädchenanteil in der Klasse“ (*AnteilWeiblich*), „Jahrgangsstufe“ (*JGS*), „Klassengröße“ (*KGröße*) sowie das Schulmerkmal „Schulform“ (*HS* und *RS*) eingesetzt (Gleichungen 7' und 11'). Damit wird geprüft, ob diese Merkmale das Ausmaß der Geschlechtsunterschiede innerhalb von Klassen beeinflussen, etwa die Jahrgangsstufe im Sinne eines Schereneffekts. Die Effekte der Klassenmerkmale wurde Der Effekt der Klassenmerkmale wurde auf Schulebene fixiert ($u_{11k}=u_{12k}=u_{13k}=0$; Gleichungen 12-14). Als Prädiktoren des schulspezifischen Einflusses des Mädchenanteils in der Klasse wurden wiederum die beiden Dummyvariablen der Schulform spezifiziert (*HS* und *RS*; Gleichung 9'), womit getestet wird ob ein etwaiger additiver Effekt der klassenspezifischen Geschlechterverteilung schulformspezifisch ausfällt.

Zur Evaluation der einzelnen Modelle wurde mit Devianztests, die eine approximativ χ^2 -verteilte Teststatistik liefern, wurde geprüft, ob Modell 1 gegenüber dem Nullmodell und

Modell 2 gegenüber Modell eine Verbesserung der Anpassung liefert (vgl. Snijders & Bosker, 1999, pp. 88-91).

Ergebnisse

Deskriptive Ergebnisse und Nullmodelle

Tabelle 1 enthält deskriptive Statistiken, die Aufteilung der Varianzen der abhängigen Variablen auf Individual-, Klassen- und Schulebene (Tests der Nullmodelle) sowie bivariate Korrelationen auf Schüler(innen)- und Klassenebene. Die durchschnittlichen Zensuren in den drei Hauptfächern lagen eng beieinander und knapp oberhalb der Note „Befriedigend (3)“. Die berichteten Leistungsaspirationen, also die Zensuren, mit denen die Schüler(innen) gerade noch zufrieden sind, liegen auffällig nah an den tatsächlichen Leistungsbewertungen. Dies spricht dafür, dass die erzielten Leistungen im Mittel nur knapp als zufriedenstellend angesehen werden. Andererseits sind die Varianzen bei den Anspruchsniveaus deutlich geringer als bei den Zensuren, was bei der Interpretation der Koeffizienten der hierarchisch linearen Modellierung zu beachten ist.

Bei allen abhängigen Variablen Maßen waren signifikante Unterschiede zwischen Klassen und Schulen zu beobachten. Bei den Leistungsbewertungen waren zwischen 13% und 21% der Varianz auf den beiden höheren Ebenen angesiedelt, bei den Leistungsaspirationen zwischen 8% und 10%. Da die Varianz innerhalb von Klassen typischerweise weit größer ist als die Varianz zwischen Klassen – Snijders und Bosker (1999) nennen 5% bis 20% als typische Varianzanteile –, ist diese Variabilität auf Ebene der Klassen und Schulen nicht nur statistisch, sondern auch praktisch bedeutsam (vgl. Hosenfeld, Helmke, Ridder & Schrader, 2001). Innerhalb dieser Varianz auf den beiden höheren Ebenen haben die Unterschiede zwischen Schulen bei den Leistungsbewertungen insgesamt eine relativ größere Bedeutung als bei den Leistungsaspirationen (5% bis 12% im Vergleich zu 2% bis 3%).

Die bivariaten Korrelationen zeigen, dass die abhängigen Variablen für die drei einbezogenen Schulfächer allenfalls moderat korreliert sind, was eine getrennte Behandlung der drei Domänen rechtfertigt. Etwas höher fallen die Zusammenhänge zwischen Leistungsbewertungen und Leistungsaspirationen aus, zumindest auf Individualebene. Kongruent zu klassischen Motivationskonzepten steht die Anspruchsniveausetzung damit in

enger Dependenz zu den erzielten Leistungen (z.B. Atkinson & Litwin, 1960; Lewin et al., 1944).

Geschlechtsunterschiede nach Kontrolle von Schulform und Jahrgangsstufe

Tabelle 2 liefert einen Überblick der Analyseergebnisse zu Modell 1. Für alle abhängigen Variablen lieferte dieses Modell eine deutlich bessere Anpassung an die Daten als das Nullmodell ($\chi^2(12) > 130.6$; $p < .001$).

Leistungsbewertungen. Nach Kontrolle von Jahrgangsstufe und Schulform waren in allen drei untersuchten Hauptfächern bedeutende Geschlechtsunterschiede bei den Leistungsbewertungen (Zensuren) zu verzeichnen: In Mathematik erhielten Jungen Zensuren, die um $\gamma_{100} = 0.146$ Notenstufen besser waren als jene der Mädchen (entspricht $d = 0.16$ Standardabweichungen). In den beiden sprachlichen Fächern erzielten dagegen Mädchen die besseren Zensuren. Besonders stark fiel der Unterschied im Fach Deutsch mit einem Unterschied von $\gamma_{100} = -.363$ Notenstufen ($d = -0.42$) aus, aber auch in der ersten Fremdsprache erzielten Mädchen Leistungen, die mit $\gamma_{100} = -.152$ Notenstufen besser bewertet wurden als jene der Jungen ($d = .17$).

Unabhängig von diesen Geschlechtseffekten innerhalb von Klassen zeigten sich in den beiden sprachlichen Fächern kleine Effekte der Geschlechterverteilung zwischen Klassen: Je höher der Mädchenanteil in den Klassen ist, desto besser wurden die Leistungen von allen Schüler(inne)n bewertet. Approximiert auf den Vorsprung von Klassen mit 75% Mädchenanteil gegenüber Klassen mit einem Mädchenanteil von 25% (entspricht einer Standardabweichung des Mädchenanteils) betrug dieser additive und um Einflüsse der individuellen Geschlechtsvariable bereinigte Effekt in Deutsch $-0.314 \cdot 0.50 = -0.157$ Notenstufen ($d = -.18$) und in der Fremdsprache $-0.305 \cdot 0.50 = -0.153$ Notenstufen ($d = -.17$). Im Fach Mathematik war ein vergleichbarer Effekt nicht nachweisbar.

Neben Geschlechtseffekten zeigte sich konsistent für alle drei Hauptfächer, dass die Leistungen von Hauptschüler(inne)n und Realschüler(inne)n mit schlechteren Zensuren bewertet werden als die Leistungen von Gymnasiast(inn)en. Dem Augenschein nach gilt dies in stärkerem Maße für Hauptschulen ($d = 0.41-0.76$) als für Realschulen ($d = 0.31-0.50$). Im Mathematik- und Fremdsprachenunterricht werden zudem mit ansteigender Jahrgangsstufe zunehmend schlechtere Zensuren vergeben.

Zentral für die verfolgte Fragestellung sind die Varianzen des Regressionskoeffizienten der Geschlechtsvariable (p_{1jk} ; vgl. Gleichung 5) zwischen Klassen und Schulen (R_1 und U_{10}). Diese indizieren, wie stark sich Klassen und Schulen im Ausmaß der Geschlechtsunterschiede bei der Leistungsbewertung unterscheiden. Die Ergebnisse erbrachten für alle drei Hauptfächer signifikante Varianzkoeffizienten R_1 , die darauf verweisen, dass sich die untersuchten Klassen darin unterscheiden, wie bedeutsam in ihnen die Geschlechtsvariable für die Leistungsbewertung ist. Zur weiteren Aufklärung dieser Klassenunterschiede wurde mit post hoc durchgeführten Residualanalysen für jede Klasse separat das Regressionsgewicht der Geschlechtsvariable p_{1jk} bestimmt und grafisch veranschaulicht (Abbildung 1). Für alle drei Hauptfächer verweisen diese Koeffizienten sowohl auf Klassen, in denen Geschlechtsunterschiede sehr stark ausgeprägt sind, als auch auf Klassen, in denen die Geschlechtsvariable eine sehr geringe oder keine Bedeutung hat. Bei Mathematik- und Fremdsprachenzensuren ist daneben auch eine substantielle Anzahl von Klassen erkennbar, in denen die Geschlechtsunterschiede in umgekehrter Richtung ausfallen als dies durch den mittleren Regressionskoeffizienten g_{100} indiziert wird, also Mädchen in Mathematik und Jungen in der Fremdsprache bessere Zensuren erzielten (Mathematik: 4% der Klassen; Deutsch: 1%; Fremdsprache: 8%). Im Gegensatz zu den deutlichen Klassenunterschieden erbrachte die Analyse von Modell 1 keine signifikante Varianz zwischen Schulen in Bezug auf die Bedeutung der individuellen Geschlechtsvariable (U_{10}). Auch für den Einfluss des Mädchenanteils zeigten sich keine statistisch bedeutsame Variabilität zwischen Schulen (U_{01}).

Leistungsaspirationen. Beim Anspruchsniveau zeigte sich nach Kontrolle der Jahrgangsstufen- und Schulformeinflüsse ebenfalls deutliche Geschlechtsunterschiede (vgl. Tabelle 2). Im Vergleich zu den Leistungsbewertungen fallen diese in allen drei Hauptfächern ungünstiger für die Mädchen aus: Durchschnittlich waren Jungen gerade noch mit Mathematiknoten zufrieden, die $\gamma_{100}=0.227$ Stufen besser lagen als jene, mit denen Mädchen gerade noch zufrieden waren. Der Unterschied von $d=0.33$ fällt hier erkennbar größer aus als bei den tatsächlich erhaltenen Noten. In Deutsch und der Fremdsprache verfolgten Mädchen zwar symmetrisch zu den Leistungsbewertungen höhere Aspirationen als Jungen, die Unterschiede indizieren aber deutlich schwächere Vorteile der Mädchen als bei diesen (Deutsch: $d=-0.29$. Fremdsprache: $d=-0.09$).

Neben diesen Geschlechtsunterschieden innerhalb von Klassen war im Fach Mathematik ein davon unabhängiger aber recht kleiner Effekt der Geschlechterverteilung zwischen Klassen zu beobachten: Je höher der Mädchenanteil in den Klassen ist, desto

niedriger waren die Mathematikaspirationen aller Schüler(innen) der Klasse. Auf den Unterschied zwischen Klassen mit 25% und 75% Mädchenanteil approximiert betrug dieser Effekt $0.165 \cdot 0.50 = 0.083$ Stufen der für die Schüler(innen) gerade noch zufriedenstellenden Mathematiknote ($d=.12$). Vergleichbare Einflüsse der Geschlechterverteilung waren bei der Anspruchsniveausetzung in Deutsch und der Fremdsprache nicht nachweisbar.

Ebenso wie bei den Leistungsbewertungen erbrachte die Analyse für alle drei Hauptfächer signifikante Varianzkoeffizienten R_1 , die auf Klassenunterschiede in der Bedeutung der Geschlechtsvariable hinweisen. Auch das Ausmaß der Geschlechtsunterschiede bei Leistungsaspirationen ist nicht in allen Klassen gleichermaßen ausgeprägt. Die post hoc durchgeführten Residualanalysen (Abbildung 2) illustrieren den Unterschied zu den Leistungsbewertungen: In allen drei Hauptfächer gibt es zwar Klassen, in denen die Geschlechtsunterschiede stärker ausgeprägt sind, und Klassen, in denen die Unterschiede weniger bedeutsam sind. Allerdings gibt es in Mathematik keine Klassen, in denen Mädchen im Vorteil sind. In Bezug auf die Fremdsprache haben in immerhin 21% der Klassen Jungen höhere Leistungserwartungen als Mädchen, obwohl in den Mehrzahl dieser Klassen Mädchen die besseren Noten erhalten.

Analog zu den Leistungsbewertungen war auf der Schulebene keine Variabilität im Hinblick auf die Bedeutung der individuellen Geschlechtsvariable statistisch abzusichern (U_{10}). Mit Ausnahme eines über Schulen hinweg variierenden Einflusses des Mädchenanteils auf die Anspruchsniveausetzung im Fach Mathematik, gilt dies auch für die den Varianzanteil U_{01} .

Erklärung der klassenspezifischen Geschlechtsunterschiede

Zur Erklärung von Klassen- und Schulunterschieden im Ausmaß der Geschlechtsunterschiede und der Bedeutung der Geschlechterverteilung wurde Modell 2 geprüft, das als *Slopes-as-Outcomes-Modell* das klassenspezifische Regressionsgewicht der Geschlechtsvariable sowie das schulspezifische Regressionsgewicht des Mädchenanteils als abhängige Variablen von strukturellen Merkmalen der Klasse und der Schule spezifiziert. Da die vorangegangenen Analysen übereinstimmend für alle Variablen statistisch nicht von Null unterscheidbare Varianzen U_{10} erbrachten wurde der Effekt der Geschlechtsvariable auf Schulebene fixiert ($u_{10k}=0$ und damit $U_{10}=0$; Gleichung 11'). Trotzdem wurde die Schulform in die Prüfung mit aufgenommen, da der Test zur Prüfung der ebenenspezifischen Varianz eine geringere Teststärke aufweist als der Test zur Prüfung der Regressionskoeffizienten

(Snijders & Bosker, 1999, p. 96). Die Ergebnisse der Parameterschätzungen mit Modell 2 finden sich in Tabelle 3.

Leistungsbewertungen. Für die Zensuren in den drei Hauptfächern zeigten Devianztests eine signifikante aber nur moderate Verbesserung der Anpassung gegenüber Modell 1 an ($\chi^2(4) > 14.4$; $p < .01$). Die statistisch absicherbaren Effekte auf das Ausmaß an Geschlechtsunterschieden in den beiden sprachlichen Fächern fallen recht ähnlich aus. Zunächst zeigte sich in beiden Fächern Evidenz für einen Schereneffekt, wonach sich die Geschlechterdiskrepanzen mit zunehmender Jahrgangsstufe vergrößern. Die signifikanten Koeffizienten b_{12} geben die Anzahl der Notenstufen an, um die sich das Regressionsgewicht p_{1jk} der Geschlechtsvariable pro Jahrgangsstufe verändert. Beispielsweise beträgt im fremdsprachlichen Unterricht der durchschnittliche Notenvorteil von Mädchen gegenüber Jungen in der 5. Jahrgangsstufe $p_{1jk} = -0.08$ Notenstufen und in der 10. Jahrgangsstufe $p_{1jk}' = p_{1jk} + 5b_{12} = -0.008 - 5 \cdot 0.030 = -0.24$ Notenstufen. An diesen Werten ist zu erkennen, dass der Schereneffekt eher klein ist ($d = -0.17$ für das genannte Beispiel. $d = -0.18$ analog für das Fach Deutsch). Bei der Mathematikzensur konnten keine Indizien für einen Schereneffekt gefunden werden.

Wiederum in Deutsch und der Fremdsprache, aber nicht in Mathematik, wurden Indizien für Effekte der Schulform gefunden. Diese verweisen darauf, dass an der Hauptschule in beiden Fächern und an der Realschule in Deutsch die Geschlechterdiskrepanzen innerhalb der Klassen etwas stärker ausgeprägt sind als am Gymnasium. Das Noten der Jungen in sprachlichen Fächern erscheinen im Vergleich zu jenen der Mädchen also in der Hauptschule und der Realschule besonders defizitär. Beispielsweise beträgt in Deutsch der durchschnittliche Vorteil der Mädchen am Gymnasium $g_{100} = -0.29$ Notenstufen und an der Realschule $g_{100} + g_{101} = -0.29 - 0.13 = -0.42$ Notenstufen (Unterschied entspricht $d = -0.15$. Deutsch Hauptschule: $d = -0.12$. Fremdsprache Hauptschule: $d = -0.17$).

In Bezug auf die Mathematikzensur war schließlich ein Effekt der Klassengröße zu beobachten, der darauf hindeutet, dass Leistungsbewertungen im Fach Mathematik in größeren Klassen weniger geschlechtsdiskrepanz ausfallen. Auch dieser den Nachteil von Mädchen abfedernde Effekt war mit einer durchschnittlichen Angleichung ihrer Noten an jene der Jungen von $b_{13} \cdot 10 = 0.015 \cdot 10 = 0.15$ Notenstufen pro 10 Schüler(innen) mehr in der Klasse klein (entspricht $d = 0.17$).

Weitere Einflüsse von strukturellen Klassen- und Schulmerkmalen zeigten sich nicht, auch nicht auf das schulspezifische Gewicht des Mädchenanteils. Insbesondere konnten keine

signifikante Abhängigkeit des Ausmaßes der Geschlechterdifferenzen von der Geschlechterverteilung in der Klasse. Setzt man als β -Fehler-Niveau 20% an, kann diesbezüglich die Hypothese abgesichert werden, dass keine Abhängigkeiten bestehen ($p > .27$).

Entsprechend der insgesamt eher kleinen Einflüsse wurden mit Modell 2 nur geringe Reduktionen der Varianz von Geschlechtsunterschieden zwischen Klassen gegenüber Modell 1 erzielt (zwischen 12% und 17% der Varianzen R_1 und U_{10} wurden aufgeklärt). Insbesondere weist auch Modell 2 noch bei Leistungsbewertungen in allen Fächern signifikante und substantielle Varianzen R_1 in der klassenspezifischen Ausprägung von Geschlechterdiskrepanzen auf.

Leistungsaspirationen. Modell 2 war gegenüber Modell 1 den Daten der Anspruchsniveausetzung in den beiden sprachlichen Fächern signifikant besser ($\chi^2(4) > 11.3$; $p < .05$) und in Mathematik marginal signifikant besser ($\chi^2(4) = 8.4$; $p < .08$) angepasst.

Indizien für einen Schereneffekt fanden sich bei den Leistungsaspirationen in den Fächern Mathematik und Deutsch. Die ohnehin schlechtere Mathematiknote, mit der sich Mädchen gerade noch zufrieden geben, war mit ansteigender Jahrgangsstufe zunehmend schlechter als jene, die für Jungen gerade noch im zufriedenstellenden Bereich liegt (Unterschied zwischen 5. und 10. Jahrgangsstufe: $d = 0.16$). In Deutsch war das Anspruchsniveau der Mädchen mit ansteigender Jahrgangsstufe zunehmend dem der Jungen überlegen; der Effekt war dabei von bemerkenswerter Stärke ($d = 0.32$). Im Fremdsprachenunterricht war keine Aufweitung der beobachteten Geschlechterdiskrepanzen abzusichern. Ähnlich wie bei den Leistungsbewertungen waren in Deutsch und der Fremdsprache Effekte der Schulform dahingehend zu beobachten, dass die Geschlechterdiskrepanzen zu Ungunsten der Jungen an der Hauptschule (Deutsch: $d = -0.15$. Fremdsprache: $d = -0.18$) und an der Realschule (nur Deutsch: $d = -0.12$) größer sind als am Gymnasium. Darüber hinausgehende Effekte von Klassen- und Schulmerkmalen auf die klassenspezifische Ausprägung von Geschlechtsunterschieden beim Aspirationsniveau konnten nicht abgesichert werden, insbesondere auch kein Effekt der Geschlechterverteilung in der Klasse ($p > .59$).

Wie bereits bei den tatsächlichen Zensuren wurden auch bei den Leistungsaspirationen mit Modell 2 eher geringe Reduktionen der Varianz von Geschlechtsunterschieden auf Klassenebene erzielt (zwischen 10% und 22% der Varianzen R_1 und U_{10} wurden aufgeklärt). Auch hier bleiben nach Aufnahme von strukturellen Klassen- und Schulmerkmalen

signifikante und substanzielle Varianzen R_1 bestehen, die darauf verweisen, dass die Bedingungsfaktoren dieser Variabilität vor allem bei qualitativ anderen Klassenmerkmalen als den hier analysierten zu suchen sind.

Diskussion

Anliegen der vorliegenden Arbeit war es zunächst, unter Kontrolle von Jahrgangsstufen- und Schulformeffekten eine valide Schätzung bestehender Geschlechtsunterschiede bei Leistungsbewertungen (Zensuren) in jenen drei mathematisch-naturwissenschaftlichen und sprachlichen Hauptfächern vorzunehmen, in denen Schüler(innen) ab der 5. Jahrgangsstufe unterrichtet werden. Kongruent zur bisherigen Literatur zeigte sich in Mathematik ein kleiner Vorteil von Jungen und in Deutsch ein mittlerer Vorteil von Mädchen. Die Effektstärken lagen dabei in dem Bereich, der von PISA 2003 für die mathematische Kompetenz und die Lesekompetenz bekannt ist (vgl. Zimmer et al., 2004). Für die erste Fremdsprache, in den allermeisten Fällen Englisch, war ein kleiner Vorteil von Mädchen zu beobachten, dessen Effektstärke in der Größenordnung des Unterschieds in Mathematik lag.

Ein ähnliches Ergebnismuster zeigte sich bei den Leistungsaspirationen, allerdings waren die Geschlechtsunterschiede in allen drei Fächern zu Ungunsten der Mädchen verschoben, so dass die Differenz zwischen Jungen und Mädchen bei diesem Parameter der Leistungsmotivation in Mathematik größer und in den beiden sprachlichen Fächern kleiner ist als bei den Leistungsbewertungen. Offensichtlich setzen sich Mädchen fachunabhängig ein niedrigeres Anspruchsniveau als Jungen. Dieser Befund fügt sich gut in die bestehende Befundlage zu Geschlechtsunterschieden im motivationalen Bereich ein, wonach Mädchen ihre Fähigkeiten und Leistungen unabhängig von deren Niveau niedriger einschätzen als Jungen und eine generell geringere Erfolgserwartung aufweisen (z.B. Ziegler, Heller & Broome, 1996).

Für Leistungsbewertungen in Deutsch und der ersten Fremdsprache erbrachten unsere Analysen Evidenz für einen kleinen aber substanziellen temporalen Schereneffekt wonach der Vorsprung der Mädchen in diesen Fächern mit ansteigender Jahrgangsstufe zunimmt. Für die Leistungsaspirationen im Fach Deutsch wurde ein analoger Effekt gefunden. Diese Ergebnisse liefern Anhaltspunkte dafür, dass die besonders bei Jungen anzutreffenden Defizite in den für diesen Fächer erforderlichen Kompetenzen (z.B. Kompetenzen im Bereich des Lesens, der Sprachproduktion oder der Textinterpretation) kumulieren und dass eine

Förderung möglichst frühzeitig ansetzen sollte. Den oft vermuteten Schereneffekt im Fach Mathematik konnten wir dagegen nur für das Anspruchsniveau bestätigen, nicht aber für die Leistungsbewertungen. Die Leistungen von Mädchen werden in diesem Fach somit von der 5. Jahrgangsstufe an konstant als schlechter bewertet.

Evidenz für einen vertikalen Schereneffekt konnten wir ebenfalls für die beiden sprachlichen Fächer erbringen. Danach sind die Geschlechtsunterschiede zum Nachteil von Jungen bei Leistungsbewertungen und –aspirationen im Deutschunterricht der Haupt- und der Realschule und im fremdsprachlichen Unterricht der Hauptschule größer als am Gymnasium. Dieses Befundmuster konvergiert gut 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 für die niedrigeren Schulformen einen besonderen Förderbedarf.

Für Leistungsbewertungen im Fach Mathematik zeigte sich eine Abhängigkeit von der Klassengröße, wonach die Benachteiligung von Mädchen in großen Klassen abgedämpft ist. Dieser Effekt war allerdings von geringer Stärke und dürfte deshalb in der Praxis wenig von Bedeutung sein. Erklärt werden kann er möglicherweise dadurch, dass Lehrkräfte in größeren Klassen ein höheres Ausmaß an Stress und kognitiver Belastung erleben (Schrader, Helmke, Hosenfeld & Ridder, 2001) und sie in der Folge bei der Leistungsbewertung in geringerem Maße Geschlechterstereotypen und Erwartungen aktivieren. Diese Erklärung ist allerdings eher vage und bedarf ebenso wie der Effekt der Klassengröße selbst der Bestätigung durch zukünftige Forschung.

Bemerkenswert ist, dass für alle drei Fächer mit recht großer Sicherheit ausgeschlossen werden kann, dass das Ausmaß der Geschlechtsunterschiede im Zusammenhang mit der Geschlechterverteilung in der Klasse steht. Ausgeschlossen werden kann damit beispielsweise, dass Mädchen in Mathematik dann eine bessere Förderung erhalten und sie weniger benachteiligt sind, wenn sie in ihrer Klasse in der Überzahl sind. Ebenso können damit bei der Bewertung von Leistungen durch Lehrkräfte Mildeeffekte gegenüber der unterzähligen Geschlechtergruppe und eine normative Funktion der Leistungen der überzähligen Geschlechtergruppe in der Klasse ausgeschlossen werden. Allerdings ist zu betonen, dass diese Ergebnisse nicht ohne weiteres auf den Fall exklusiver Mädchen- oder Jungensklassen übertragen werden können, einerseits weil unsere Stichprobe keine monoedukativen Klassen umfasste, andererseits weil die monoedukative Unterrichtung bestimmte Erwartungen bei Schüler(inne)n und Lehrkräften triggert, die bei gemischter Geschlechterverteilung nicht zum Tragen kommen (Ziegler, Broome & Heller, 1998).

Unabhängig von den auszuschließenden Einflüssen auf die Größe der Diskrepanz zwischen den Geschlechtern hatte die Geschlechterverteilung einen gleichförmigen aber schwachen Einfluss auf die Leistungsbewertungen aller Schüler(innen) in Deutsch und der ersten Fremdsprache sowie Leistungsaspirationen aller Schüler(innen) im Fach Mathematik: Je höher der Mädchenanteil in der Klasse war, desto besser waren die Zensuren von Mädchen und Jungen in Deutsch und der ersten Fremdsprache und desto geringer war das Anspruchsniveau von Mädchen und Jungen in Mathematik. Zur Erklärung dieser monotonen Effekte der Geschlechterverteilung in der Klasse sind zwei Interpretationsstränge denkbar: Einerseits könnte die Minorität von der Majorität Wissen und Einstellungen übernehmen, mithin von ihr lernen. Beispielsweise könnten nach dieser auf Lernprozesse abhebenden Interpretation Jungen in einer Klasse mit hohem Anteil an Mädchen deren niedrigeres Anspruchsniveau im Fach Mathematik übernehmen. Andererseits können zumindest die Effekte bei den Zensuren auch als Erwartungseffekte bei der Bewertung durch Lehrkräfte interpretiert werden. So ist es plausibel, dass Lehrkräfte neben Erwartungen in Bezug auf einzelne Schüler(innen) auch Erwartungen und Einschätzungen auf Klassenebene aufweisen und Klassen beispielsweise als „mathematisch gute Klasse“ oder „sprachlich schlechte Klasse“ betrachten. Mit höherer Wahrscheinlichkeit dürften Lehrkräfte Klassen mit vielen Mädchen aufgrund deren besseren Leistungen als „sprachlich gute Klasse“ ansehen, wobei dieses Schema bei der Bewertung der Leistungen aller Schüler(innen), also auch der in Unterzahl vertretenen Jungen, im Sinne eines Erwartungseffekts wirksam werden könnte. Die Aufklärung, zu welchem Anteil diese beiden Prozesse für den gleichförmigen Effekt der Geschlechterverteilung auf die Leistungsbewertungen in sprachlichen Fächern und die Anspruchsniveausetzung in Mathematik verantwortlich sind, dürfte zu einem besseren Verständnis der Prozesse führen, die für Entwicklung und Aufrechterhaltung von Geschlechterdiskrepanzen im schulischen Unterricht verantwortlich sind und stellt eine interessante Aufgabe zukünftiger Forschung dar.

Eine zentrale Frage, die wir mit der vorliegenden Arbeit zu beantworten suchten, richtete sich auf die Variabilität von Geschlechtsunterschieden zwischen Klassen und Schulen. Aufgrund der multiplen Quellen geschlechtsspezifischer Sozialisation im Klassenraum - Lehrkräfte, Mitschüler(innen) und Unterrichtsmedien - nahmen wir an, dass sowohl Klassen und Schulen existieren, die sich durch starke Geschlechtsunterschiede charakterisieren lassen, als auch solche, in denen Geschlechtsunterschiede nicht oder nur abgeschwächt zum Tragen kommen. Für die Klassenebene konnte diese Annahme bestätigt werden, wobei die Variabilität zwischen Klassen insbesondere in Bezug auf

Leistungsbewertungen substanziell war, aber auch in Bezug auf Leistungsaspirationen gegenüber Zufallsschwankungen statistisch abgesichert werden konnte, die beispielsweise durch den „Stichprobenfehler“ bei der Zusammenstellung von Klassen begrenzter Größe aus den Populationen männlicher und weiblicher Schüler(innen) bedingt sein können. Diese Variabilität zwischen Klassen konnte durch die bereits diskutierten Einflüsse von Jahrgangsstufe, Schulform und anderen strukturellen Merkmalen von Klasse und Schule nur in geringem Maße aufgeklärt werden (zwischen 10% und 22% Varianzaufklärung) und war auch nach Kontrolle dieser Merkmale noch substanziell. Post-hoc durchgeführte Residualanalysen zeigten für Leistungsbewertungen und Leistungsaspirationen 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 sehr schwach zum Tragen kommt. In diesen Klassen, die durchaus 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 des gängigen Stereotyps verteilt und damit besonders ungleich sind. Als entscheidende Ebene für diese Variabilität hat sich in unserer Analyse die Klassenebene erwiesen; für eine analoge Variabilität zwischen Schulen fanden wir keine Indizien. 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.

Diese Variabilität im Ausmaß, in dem Geschlechtsunterschiede zum Tragen kommen, kann als weiteres Argument dafür gewertet werden, dass die Ursachen von Geschlechtsunterschieden im Leistungs- und Motivationsbereich primär in einer geschlechtsspezifischen Sozialisation und weniger in anlagebedingten Unterschieden zu suchen sind (z.B. Halpern, 2000; Trautner, 1994). 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 (allen voran die Eltern; vgl. Dresel et al., 2001) der Klassenraum 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. dass Jungen begabter für mathematisch-naturwissenschaftliche Domänen seien; Heller, et al.,

2000; Rustemeyer, 1999) und Interaktionen mit Schüler(inne)n häufig in geschlechtsspezifischer Weise initiieren (z.B. interagieren sie mit Jungen generell häufiger als mit Mädchen; Jones & Dindia, 2004). Daneben wurde Evidenz dafür erbracht, dass Jungen eher von einem konkurrenzorientierter und Mädchen eher von einem kooperationsorientierten Klassenklima profitieren (zsf. Beerman et al., 1992). Ein fruchtbares Feld zukünftiger Forschung ist die Überprüfung, ob diese durchschnittlich geschlechtsspezifischen 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 Variabilität zwischen Klassen auf eine Variabilität im Ausmaß von Erwartungseffekten bei der Leistungsbewertung durch Lehrkräfte und auf eine Variabilität in den tatsächlichen Leistungen zurückgeht. Ergebnisse von standardisierter 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 Ergebnissen zu den Zensuren der Schüler(innen) 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) und für zukünftige Wahlentscheidungen kaum zu unterschätzen sein.

Neben den Lehrkräfteinflüssen dürfte aber auch der Geschlechtsrollenkonformität von Einstellungen, Überzeugungen und Verhaltensweisen von Mitschüler(inne)n bedeutender Einfluss zukommen. Dafür sprechen beispielsweise die Befunde von Finsterwald und Dresel (2005), die für Klassen der vierten Jahrgangsstufe belegen konnten, dass für die Wahrnehmung von Lern- und Leistungszielen in diesen Klassenräumen (als wichtiger Aspekt des Klassenklimas) die Ziele von Mitschüler(inne)n wenigstens so bedeutsamer wie jene, die Lehrkräfte verfolgen. Zur Illustration der Evidenz geschlechtsspezifischer Wertevorstellungen im Klassenraum mag ein anekdotischer Bericht aus einer siebten Klasse dienen, in der Jungen (mit Autorisierung durch die Klassenlehrkraft!) diskriminierende Photographien mit leicht bekleideten und auf Motorhauben sitzenden Frauen im Klassenzimmer aufhängten. Ganz offensichtlich dürften in dieser Klasse traditionelle und abwertende Geschlechtsrollenvorstellungen in besonderer Weise aktiviert sein. In anderen Klassen muss dies nicht der Fall sein. Durchaus plausibel ist es also, dass sich die Schüler(innen)

unterschiedlicher Klassen darin unterscheiden, wie sehr sie in ihren Wertevorstellungen und Idealselbstbildern den traditionellen Geschlechterstereotypen verpflichtet sind, sie es etwa als „unweiblich“ betrachten, in Mathematik gut zu sein, und als „unmännlich“ ansehen, in sprachlichen Fächern gute Leistungen zu erzielen. Hierbei dürften komplexe Prozesse der Einstellungs- und Werteübernahme eine Rolle spielen, die je nach Klasse in unterschiedlich geschlechtsspezifischen Überzeugungen resultieren. Zur Aufklärung der Variabilität im Ausmaß von Geschlechtsunterschieden scheint ein genauer Blick auf dieses Interaktionsgeschehen lohnend, vor allem in Klassenräumen, die als besonders geschlechtsegalitär oder als besonders geschlechtsdiskriminativ gelten können.

Als pädagogische Ansatzstellen impliziert die von uns identifizierte Variabilität schließlich, dass Lehrkräfte stärker als bisher für mögliche Chancendisparitäten sensibilisiert, über die Ursachen von Geschlechtsunterschieden aufgeklärt und sie systematisch 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 den traditionellen Rollenvorstellungen widersprechen, als geeignet.

Literatur

- Atkinson, J. W. & Litwin, G. H. (1960). Achievement motive and test anxiety conceived as motive to approach success and motive to avoid failure. *Journal of Abnormal and Social Psychology, 33*, 52-63.
- 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.
- Broome, P. (1998). *Implizite Begabungstheorien und erlernte Hilflosigkeit*. Frankfurt/M., Germany: Lang.
- 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.
- Dresel, M., Schober, B., Stöger, H. & Ziegler, A. (2005). *Die Ulmer Motivationstestbatterie UMTB: Konzeption und erste Ergebnisse der Validierung*. Vortrag auf der 9. Tagung der Fachgruppe Pädagogische Psychologie der Deutschen Gesellschaft für Psychologie in Halle.
- 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. (2005). Geschlechtsrollensstereotypen 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. Manuskript in Druck.
- Finsterwald, M. & Dresel, M. (2005). *Wahrgenommene Zielorientierungen in Grundschulklassen: Schüler(innen)- und Klassenraumeffekte*. Manuskript eingereicht.
- Halpern, D. F. (2000). *Sex differences in cognitive abilities* (3rd ed.). Mahwah, NJ: 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.
- 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.
- 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., 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.
- 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.
- Marshall, S. P., & Smith, J. D. (1987). Sex differences in learning mathematics: A longitudinal study with item and error analyses. *Journal of Educational Psychology*, 79, 372-383.
- 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.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F. & Congdon, R. T. (2001). *HLM 5* [Computer Software, Version 5.04]. Chicago, IL: 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.
- 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.
- Stipek, D. J. & Gralinski, J. H. (1991). Gender differences in children's achievement-related beliefs and emotional responses to success and failure in mathematics. *Journal of Educational Psychology*, 83, 361-371.
- 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* (Enzyklopädie der Psychologie Serie I, Band 1, S. 167-195). Göttingen: Hogrefe.
- 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., Heller, K. A. & Broome, P. (1996). Motivational preconditions of gifted and highly gifted girls in physics. *High Ability Studies*, 7, 129-143.
- 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.

Fußnote

¹ Klassen mit weniger als 10 Schüler(inne)n wurden vor der Stichprobenszusammenstellung ausgeschlossen.

Tabelle 1. Deskriptive Statistiken, Varianzverteilung auf Individual-, Klassen- und Schulebene sowie bivariate Korrelationen auf Individualebene (oberhalb der Diagonale) und Klassenebene (unterhalb der Diagonale)

			Varianzaufteilung ^a			Bivariate Korrelationen ^b						
	<i>M</i>	<i>SD</i>	Innerhalb Klassen	Zwischen Klassen	Zwischen Schulen	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Zensur Mathematik	2.784	0.937	0.764	0.068***	0.044***	-	.352	.385	.706	.289	.319	.063
(2) Zensur Deutsch	2.838	0.864	0.599	0.062***	0.094***	.526	-	.485	.337	.777	.373	-.232
(3) Zensur Fremdsprache	2.805	0.914	0.722	0.071***	0.046***	.516	.513	-	.295	.304	.707	-.103
(4) Aspirationen Mathematik	2.960	0.685	0.431	0.030***	0.009***	.439	-.021	.164	-	.468	.453	-.067
(5) Aspirationen Deutsch	3.076	0.574	0.303	0.019***	0.008***	.206	.321	.299	.528	-	.451	-.145
(6) Aspirationen Fremdsprache	2.997	0.661	0.398	0.030***	0.012***	.197	.074	.572	.510	.606	-	-.086
(7) Geschlecht ^c	0.514	0.500	-	-	-	-.046	-.249	-.205	.143	-.060	.006	-
(8) Jahrgangsstufe ^d	7.213	1.606	-	-	-	.297	-.040	.293	.364	.194	.338	.041

Anmerkungen: $I=9.207$ Schüler(innen) in $J=498$ Klassen. Leistungsbewertungen und Leistungsaspirationen sind auf der üblichen Notenskala skaliert. ^a Für Klassen- und Schulebene ist das Signifikanzniveau des Test daraufhin angegeben, ob dieser Varianzanteil größer Null ist. ^b Alle dargestellten Korrelationskoeffizienten spiegeln Pearson-Korrelationen wieder, außer jenen mit dem Geschlecht (Spearman-Rang-Korrelationen). Alle $|r|>.02$ auf Individualebene und alle $|r|>.09$ auf Klassenebene: $p<.05$. ^c 0=männlich. 1=weiblich. ^d Mittelwert und Standardabweichung auf Klassenebene.

*** $p<.001$.

Table 2. Hierarchisch lineare Modellierung von Klassen- und Schuleffekte unter Berücksichtigung von Geschlecht, Jahrgangsstufe, Klassengröße und Schulart (Modell 1)

	Zensuren			Aspirationen			df
	Mathematik	Deutsch	Fremdsprache	Mathematik	Deutsch	Fremdsprache	
Schätzung der Regressionskoeffizienten							
Achsenabschnitt (g_{000})	2.565***	2.507***	2.587***	2.977***	3.045***	2.999***	51
Weiblich (g_{100})	0.146***	-0.363***	-0.152***	0.227***	-0.165***	-0.060**	53
Anteil Weiblich (g_{010})	0.077	-0.314***	-0.305**	0.165**	-0.080	-0.005	53
Jahrgangsstufe ($b_{02}=g_{020}$)	0.075***	-0.007	0.079***	0.058***	0.027***	0.055***	495
Hauptschule (g_{001})	0.449***	0.657***	0.377***	-0.117**	-0.022	-0.106**	51
Realschule (g_{002})	0.294***	0.435***	0.317***	0.037	0.087**	0.079**	51
Schätzung der Varianzanteile							
Residualvarianz innerhalb Klassen (E)	0.750	0.552	0.704	0.415	0.291	0.390	-
Residualvarianz zw. Klassen (R_0)	0.053***	0.063***	0.053***	0.021***	0.017***	0.021***	389
Residualvarianz zw. Schulen (U_{00})	0.010***	0.018***	0.022***	0.005***	0.006***	0.009***	48
Einfluss Geschlecht zw. Klassen (R_1)	0.035***	0.062***	0.046***	0.016***	0.022***	0.028***	444
Einfluss Geschlecht zw. Schulen (U_{10})	0.001	0.002	0.004	0.000	0.001	0.002	50
Einfluss Anteil Weiblich zw. Schulen (U_{01})	0.004	0.019	0.075	0.024**	0.002	0.042	50

Anmerkungen: $I=9.207$ Schüler(innen) in $J=498$ Klassen in $K=54$ Schulen. Leistungsbewertungen und Leistungsaspirationen sind auf der üblichen Notenskala skaliert. Die Variable „Weiblich“ wurde am Klassenmittelwert, die Variablen „Anteil Weiblich“ und „Jahrgangsstufe“ am Schulmittelwert zentriert. Die Dummyvariablen „Hauptschule“ und „Realschule“ wurden nicht zentriert, der Achsenabschnittsparameter geben deshalb die Werte für die Referenzkategorie „Gymnasium“ wider.

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

Tabelle 3. Hierarchisch lineare Prädiktion klassenspezifischer Geschlechtseffekte (Koeffizient p_{1jk}) und schulspezifischer Effekte der Geschlechterverteilung (Koeffizient b_{01k}) durch Klassen- und Schulmerkmale (Modell 2)

	Zensuren			Aspirationen			df
	Mathematik	Deutsch	Fremdsprache	Mathematik	Deutsch	Fremdsprache	
Schätzung der Regressionskoeffizienten							
AV: Koeffizient Weiblich (p_{1jk})							
Achsenabschnitt (g_{100})	0.170***	-0.286***	-0.093***	0.252***	-0.121***	-0.023	494
Anteil Weiblich ($b_{11}=g_{110}$)	-0.148	-0.148	-0.086	0.035	-0.023	-0.049	494
Jahrgangsstufe ($b_{12}=g_{120}$)	0.005	-0.031**	-0.030**	0.022**	-0.037***	-0.013	494
Klassengröße ($b_{13}=g_{130}$)	-0.015**	-0.006	-0.003	-0.004	0.000	0.002	494
Hauptschule (g_{101})	0.045	-0.107**	-0.151**	-0.017	-0.084**	-0.118**	494
Realschule (g_{102})	-0.078	-0.131**	-0.046	-0.045	-0.071**	-0.012	494
AV: Koeffizient Anteil Weiblich (b_{01k})							
Achsenabschnitt (g_{010})	-0.014	-0.427**	-0.448**	0.218**	-0.104**	-0.049	51
Hauptschule (g_{011})	0.096	0.125	0.154	-0.145	-0.050	0.013	51
Realschule (g_{012})	0.291	0.357	0.380	0.019	0.238	0.253	51
Schätzung der Varianzanteile							
Einfluss Geschlecht zw. Klassen (R_1)	0.030**	0.057***	0.044***	0.014**	0.018***	0.027***	494
Einfluss Anteil Weiblich zw. Schulen (U_{01})	0.003	0.008	0.018	0.019**	0.001	0.040	48

Anmerkungen: $I=9.207$ Schüler(innen) in $J=498$ Klassen in $K=54$ Schulen. Aufgeführt sind nur Koeffizienten, die sich gegenüber Modell 1 ändern konnten (vgl. Tabelle 2). Leistungsbewertungen und Leistungsaspirationen sind auf der üblichen Notenskala skaliert. Die Variable „Weiblich“ wurde am Klassenmittelwert, die Variablen „Anteil Weiblich“ und „Jahrgangsstufe“ am Schulmittelwert zentriert. Die Dummyvariablen „Hauptschule“ und „Realschule“ wurden nicht zentriert, die Achsenabschnittsparameter geben deshalb die Werte für die Referenzkategorie „Gymnasium“ wider.

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

Abbildungsbeschriftungen

Abbildung 1

Geschlechtsunterschiede bei Leistungsbewertungen in den einzelnen Klassen

(Regressionskoeffizienten $p_{1,jk} = g_{100} + u_{10k} + r_{1jk}$ der Geschlechtsvariable für die 497 einzelnen Schulklassen). Eine Linie entspricht etwa drei Schulklassen.

Abbildung 2

Geschlechtsunterschiede bei Leistungsaspirationen in den einzelnen Klassen

(Regressionskoeffizienten $p_{1,jk} = g_{100} + u_{10k} + r_{1jk}$ der Geschlechtsvariable für die 497 einzelnen Schulklassen). Eine Linie entspricht etwa drei Schulklassen.

Abbildung 1

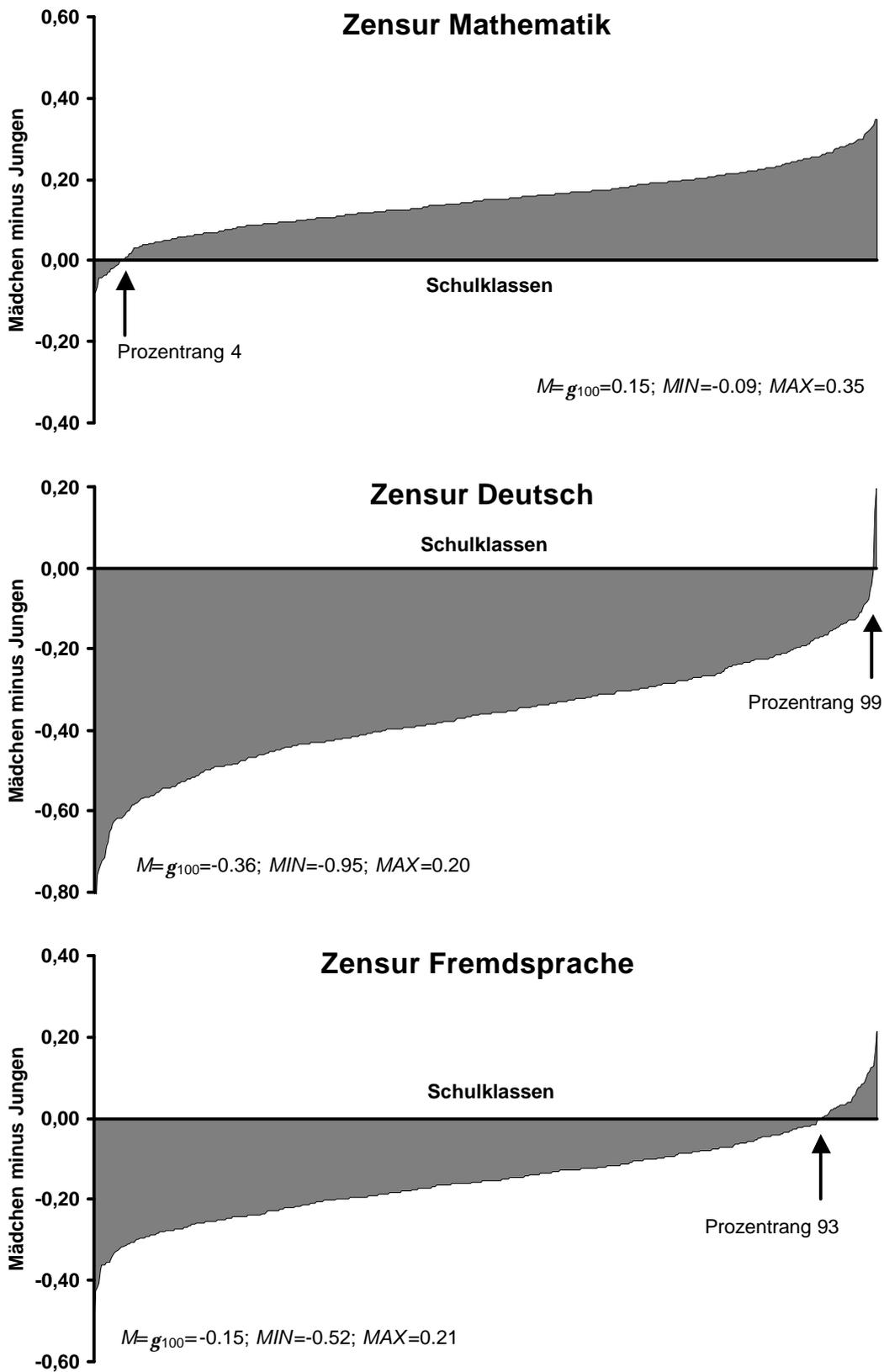


Abbildung 2

