

Peter Birkel

Geschlechtsspezifische Effekte bei der Zensurengebung im Fach Mathematik? *Gender-specific effects on marking mathematics?*

Zusammenfassung:

Anhand der Daten aus 3216 Zeugnissen der 5. und 6. Klassen von Haupt- und Realschulen im Regierungsbezirk Südwürttemberg konnte weder die von Carter (1952) berichtete Interaktion zwischen dem Geschlecht der Schülerinnen und Schüler und dem der Lehrkräfte bezüglich der Notengebung im mathematischen Bereich noch die mildere Notengebung durch Lehrerinnen bestätigt werden. Erhielten damals die Mädchen den Jungen gegenüber bessere Noten, so bekommen heute die Jungen zwar signifikant bessere Noten im Fach Mathematik, wenngleich der Vorsprung relativ gering und wenig relevant ist. Als vergleichsweise starker Einflussfaktor erwies sich die Schulartzugehörigkeit. Bei einer Varianzklärung von 5% erhielten die Realschülerinnen und -schüler bessere Noten als die Hauptschülerinnen und -schüler.

Schlüsselbegriffe: Zeugnisnoten, Schülergeschlecht, Lehrergeschlecht, Schultarteffekt

Summary:

The results of a study based on grades from 3216 school reports of 5th- and 6th-graders in secondary schools (Haupt- and Realschule) confirmed neither the interaction effect of the gender of pupils and the gender of teachers on grade assessment in mathematics found by Carter (1952) nor the awarding of better marks by female teachers. Then girls obtained far better marks, but today boys obtain significantly better marks than the girls, although the gap in performance levels remains relatively small. A strong influencing factor was found to be the type of school pupils attend. With an explained variance of 5% pupils from Realschulen obtained better marks than those from Hauptschulen.

Keywords: school grades, gender of pupils, gender of teacher, effect of type of school

1. Untersuchungsanlass

Im Rahmen einer Lehrveranstaltung zu Problemen der Leistungsbeurteilung von Schülern wurde der von Ingenkamp (1995⁹) 1971 erstmals abgedruckte Beitrag von Robert Scriven Carter aus dem Jahr 1952 kritisch diskutiert. Carter fand heraus, dass sich die Jungen und Mädchen bei einem Intelligenztest und einem Algebra-Test nicht wesentlich unterschieden, auch wenn die Jungen bei letzterem im Schnitt ein wenig besser abschnitten als die Mädchen (Ingenkamp, 1995, S. 154). Trotz fast gleicher Fähigkeiten im Test schnitten die Jungen bei der Vergabe von Zensuren deutlich schlechter ab als die Mädchen ($t_{hom} = 2,80$, $df = 233$, $p < .01$, nachberechnet nach den Originalangaben). Alle Lehrkräfte gaben den Mädchen bessere Zensuren als den Jungen.

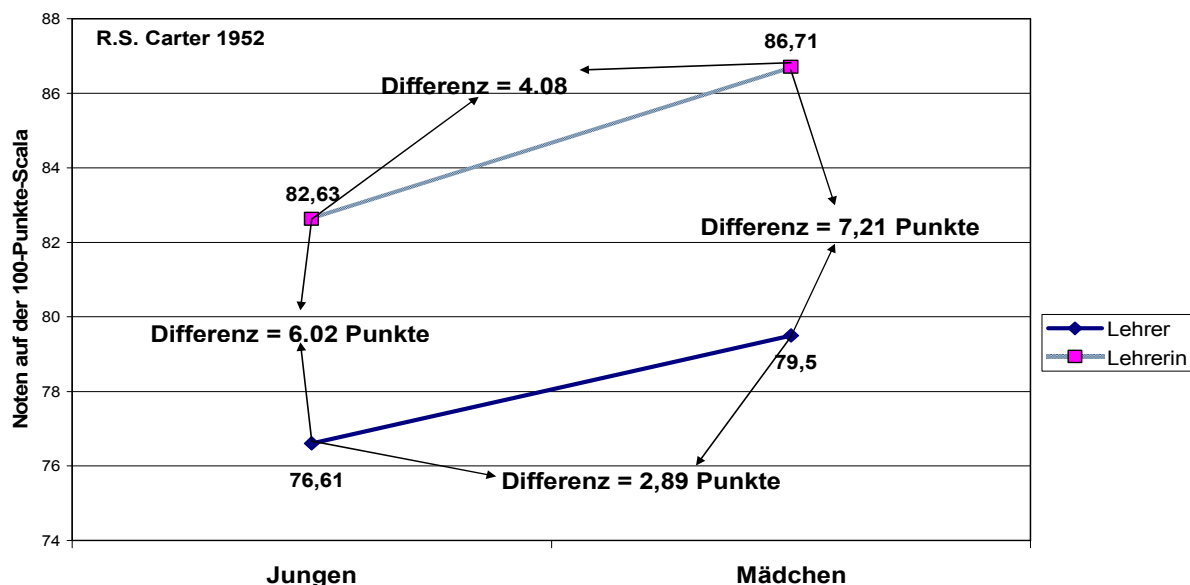


Abbildung 1: Ergebnisse der Untersuchung von Carter (1952)

Bei den Mädchen betrug der Vorteil in den Zensuren - erteilt von den Lehrern - 2,89 Punkte auf der 100-Punkte-Skala ($t_{het} = 2,25$, $df = 131$, $p < .05$). Die Lehrerinnen gaben den Mädchen einen Zensurenvorsprung von 4,08 Punkten und benachteiligten damit die Jungen noch etwas

stärker ab als ihre männlichen Kollegen ($t_{hom} = 2,27$, $df = 100$, $p < .05$), denn beim Test hatten ja Jungen und Mädchen praktisch gleichgut abgeschnitten. Trotzdem gaben sie den Jungen immerhin noch bessere Zensuren als die, die ihre männlichen Kollegen für die Mädchen erteilt hatten. Damit kann man nun fragen, ob weibliche Lehrkräfte generell mildere Zensuren erteilten als ihre männlichen Kollegen oder ob sie vielleicht auch nur andere Kriterien bei ihrer Beurteilung zugrunde legten ($Diff = 6,44$ Punkte, $t_{het} = 5,73$, $df = 196$, $p < .001$).

Diese Interaktion bei Leistungsbeurteilungen zwischen dem Geschlecht der Lehrkräfte und dem der Schülerinnen und Schüler entspricht in etwa der „matching“-Hypothese, die Marsh, Martin & Cheng (2008) in Bezug auf die Motivierung von Schülerinnen und Schülern überprüfen wollten, die von Lehrerinnen und Lehrern unterrichtet wurden. Bei gleichem Geschlecht von Lehrkräften und Schülerinnen und Schülern könnten die Lernenden profitieren, bei unterschiedlichem Geschlecht würden sich eher negative Effekte ergeben. Zu erwarten, dass die Lernenden dann auch bei geschlechtsgleichen Lehrern bessere Zensuren erhalten, erschiene dann als naheliegend, denn die besseren Leistungen müssten sich auch in besseren Zensuren niederschlagen. Demgegenüber wurde die „invariance“-Hypothese formuliert, die davon ausgeht, dass die Motivation der Schülerinnen und Schüler höchstens marginal vom Geschlecht der Lehrkraft abhängt. In diesem Falle würden sich vermutlich auch die Leistungsbeurteilungen kaum unterscheiden. In den Ergebnissen von Marsh et al. (2008) zeigte sich ein überraschend geringer Anteil der Varianz als abhängig vom Geschlecht der Lehrkraft, dem Geschlecht der Schülerinnen und Schüler und von der Interaktion zwischen den Geschlechtern. Ihre Ergebnisse unterstützen also eher die Invarianzhypothese.

Da an einer Pädagogischen Hochschule der Anteil weiblicher Studierender mit geschätzten 85% traditionell sehr hoch liegt, entzündete sich eine heftige Diskussion an der Frage, ob die Untersuchung von Carter (1952), die 60 Jahre zurück liegt, noch die Verhältnisse an heutigen Schulen valide widerspiegeln könne. Die *gender*-Problematik stand damals sicher noch nicht so sehr im Zentrum der Überlegungen zu dieser Untersuchung, obwohl mit der Interaktion zwischen dem Geschlecht der Lehrkräfte und dem der Schülerinnen und Schüler klar wurde, dass bei Beurteilungsprozessen in der Schule die Geschlechtszugehörigkeit von beurteilenden Lehrkräften und beurteilten Schülerinnen und Schülern vielleicht doch eine Rolle spielen konnte. Hannover (eingereicht) kommt zu dem Resümee, dass eine abschließende Bewertung der Frage, ob das Geschlecht einer Lehrkraft sich auf die Entwicklung des Lernenden auswirkt, nicht möglich ist. Man wird wohl davon ausgehen dürfen, dass Zeugniszensuren die weitere Entwicklung der Schülerinnen und Schüler beeinflussen. Ob sich aber das Geschlecht der Lehrkraft auch heute noch auf die Beurteilung von Schülerinnen und Schülern überhaupt auswirkt, sollte nach der Auffassung der Studierenden in einer Untersuchung unter heutigen Bedingungen erneut geprüft werden, zumal Klein (2004) in seiner Untersuchung keine solche Interaktion findet. Er berichtet allerdings auch über die milderen Zensuren, die von Lehrerinnen erteilt werden.

Im Hinblick auf die Beurteilungsgerechtigkeit der Lehrkräfte bezweifelten vor allem die Studentinnen, dass auch heute noch die Lehrerinnen andere Zensuren erteilen würden im Vergleich zu ihren männlichen Kollegen. Ob man die Zensuren der weiblichen Lehrkräfte dabei als mild oder die männlichen als zu streng ansehen müsste, ließ sich nicht entscheiden. Ein wichtiger Grund für diese Zweifel war vor allem die Vermutung, dass es doch im Fach Mathematik – wie in keinem anderen Fach sonst – möglich sein sollte, Schülerleistungen einigermaßen objektiv, d.h. unabhängig von der bewertenden Person, zu beurteilen. Wenn eine objektiv feststellbare Anzahl von Fehlern vorliege, dürfte das Geschlecht der beurteilenden Lehrkräfte kaum eine Rolle spielen. Die Untersuchung von Birkel (2005) war den Studierenden noch nicht bekannt. Man interpretierte die bei Carter berichteten Unterschiede zwischen den Beurteilungen von Lehrern und Lehrerinnen als „Gerechtigkeitsdefizit“ bei den Lehrerinnen und damit als Vorwurf ihnen gegenüber. Die von Carter (1952) aufgeführten Belege aus früheren Zeiten (Garner 1935, Day 1938, Douglass 1937, Shinnerer 1944, Edminston 1943) erbrachten zwar ebenfalls Hinweise darauf, dass Lehrerinnen nicht nur mildere Zensuren erteilten, sondern auch dass es eine Interak-

tion zwischen dem Geschlecht der Lehrkräfte und dem der Schülerinnen und Schüler geben könnte, die eine besondere Begünstigung der Mädchen durch Lehrerinnen nahelegt. Eine neuere Untersuchung von Klein (2004) in Israel scheint die generelle Tendenz der Lehrerinnen, mildere Zensuren zu geben, zu stützen. Aber hier wurde das Problem kontrovers diskutiert, inwieweit eine Untersuchung aus Israel zu übertragen sei auf deutsche Verhältnisse. Zudem wurde diskutiert, ob eine generelle Tendenz zu milderer Beurteilung – falls sie denn überhaupt nachweisbar wäre – ihren Ursprung haben könnte

- a) In der Tatsache, dass Lehrerinnen *dem weiblichen Geschlecht angehören* oder vielleicht doch eher
- b) in einem *Bemühen* von Lehrerinnen, über die milderen Zensuren die *Professionalität* ihres Unterrichts zu dokumentieren.

Nachdem in den vergangenen 60 Jahren der Anteil der Lehrerinnen in den Kollegien ständig gestiegen sei und sich die Situation in Schule und Unterricht vermutlich wohl grundlegend verändert habe, müsse diese Fragestellung – so die Studierenden – dringend erneut untersucht werden. Mit einer Bestätigung der Beurteilungsunterschiede zwischen Lehrerinnen und Lehrern wäre man bereit, diese eher als verursacht durch die geschlechtsspezifische Sozialisation, die zur Anwendung unterschiedlicher Gewichtung von Urteilkriterien führt, und damit als immer wieder beobachtbar zu betrachten. Die eventuell unterschiedlich verwendeten Kriterien müssten dann in Folgeuntersuchungen genauer identifiziert werden. Würde dieser Unterschied nicht wieder auftreten, käme als Erklärung die inzwischen weitgehend abgeschlossene Professionalisierung der Lehrerinnen in Frage, die den Wert ihrer Arbeit nicht mehr so sehr über die besseren Zensuren zu beweisen hätten, die sie erteilen.

Einen Ansatz für die Erklärung eventuell vorhandener Beurteilungsunterschiede zwischen Lehrerinnen und Lehrern liefert Flaake (1989) mit ihrer psychoanalytisch interpretierten Sozialisation der Geschlechter. Sie schreibt:

„Viele Lehrer ziehen sich auf die vermeintliche Objektivität von Leistungsbewertungen - auf allgemeine, von der Befindlichkeit und der Situation einzelner abstrahierende Prinzipien wie Gerechtigkeit - zurück, die sie entlasten vom Druck persönlicher Entscheidungen und damit auch von persönlichen Konflikten und Verantwortlichkeiten. Viele Lehrerinnen fühlen sich dagegen bei der Zensurengebung stark persönlich gefordert: sie haben die Konsequenzen, die eine schlechte Note für die Schülerinnen und Schüler hat, vor Augen, haben deshalb Probleme, solche Zensuren zu geben und suchen oft nach Wegen, sie vermeiden zu können. Ihre Einfühlung in die Situation der Schülerinnen und Schüler macht es vielen Lehrerinnen schwer, zu ihrer institutionell vorgegebenen Aufgabe der Leistungsbewertung zu stehen“ (S. 349).

Flaake transportiert damit sicher eine vielfach vermutete Erklärung der Urteilsunterschiede, liefert aber keinen empirischen Beweis für die Richtigkeit ihrer Interpretation. Von daher erscheint es als sinnvoll und notwendig, das Urteilsverhalten der Lehrkräfte unter heutigen Bedingungen neu zu untersuchen, zumal unsere Erwartungen eher im Gegensatz zu Flaakes Überlegungen stehen.

Im Gegensatz zur *gender*-Problematik der beurteilenden Lehrkräfte ist diese Problematik im Hinblick auf die Abhängigkeit der Beurteilungen vom Schülergeschlecht in den letzten Jahrzehnten immer wieder Gegenstand von empirischen Untersuchungen gewesen. Zunächst führt Carter (1952) eine Reihe von empirischen Belegen aus früheren Zeiten für die Tatsache an, dass unabhängig vom Geschlecht der urteilenden Lehrkraft Mädchen in der Sekundarstufe I generell, und damit auch bei der Vergabe von Zensuren im Fach Mathematik deutliche Vorteile gegenüber von Jungen haben (s. S. 148 -150). Diese Vorteile der Mädchen ließen sich in einer neueren Untersuchung im Bereich der Primarstufe nicht durchgängig bestätigen (Tiedemann & Faber 1994, S. 102). Diese Autoren berichten, dass in den meisten Untersuchungen, die sie gefunden hatten, kein Geschlechtsunterschied bei Grundschulern im Fach Mathematik festgestellt werden konnte (z.B. Callahan & Clements 1984, Dossey, Mullis Lindquist & Chambers 1988, McKay 1979,

Siegel 1968), einige sehen Vorteile für die Mädchen dieser Alterstufe (Brandon, Newton & Hammond 1985, Hawn, Ellet & Des Jardines 1981, Potter & Lewy 1968, Shipman 1972), wenige beschreiben Vorteile für die Jungen (z.B. Lewis & Hoover 1986). Auch Ratzka (2003) weist auf eher gleiche Leistungen bei den Grundschulern hin.

Neuere Untersuchungen in der Sekundarstufe I (Baumert et al. 1997, 2000, Lehmann & Peek 1997, Lehmann, Gänsfuß & Peek 1999, Dt. PISA-Konsortium 2001, 2004, 2007, Stürzer 2003a, Kampshoff 2007, Schreiner, Breit & Haider 2008, zusammenfassend vbw 2009) scheinen eher für die Tendenz zu sprechen, dass im Falle unterschiedlicher Leistungen die Jungen im Fach Mathematik den Mädchen eher überlegen zu sein scheinen, wenngleich durchaus spezielle Bereiche im Fach Mathematik identifizierbar seien, in denen auch die Mädchen sich den Jungen überlegen zeigen können. Die mathematisch-naturwissenschaftlichen Fächer scheinen zumindest in der Sekundarstufe eher eine Jungendomäne zu sein, während die sprachlichen und die dem musisch-ästhetischen Bereich zuzuordnenden Fächer eindeutig Mädchendomänen sind (s. z.B. Stürzer 2003a, Dt. PISA-Konsortium 2007). Die Überlegenheit der Mädchen in diesen Bereichen scheint viel größer zu sein, als die der Jungen im mathematisch-naturwissenschaftlichen Bereich, wodurch man die Mädchen auch in der Sekundarstufe als die insgesamt erfolgreicherer Schüler ansehen kann. Auch wenn der Bildungserfolg sicher nicht monokausal durch das Geschlecht erklärt werden kann, so scheint Bildungserfolg oder –misserfolg eben doch auch eine Geschlechterfrage zu sein. Gesellschaftliche Faktoren müssen hier mitberücksichtigt werden (Crotti 2006). Erklärungsversuche für Geschlechtsunterschiede bei Schülerinnen und Schülern reichen von genetischen Veranlagungen über evolutionspsychologische Hypothesen bis hin zur geschlechtsspezifischen Sozialisation und zu psycho-sozialen Ansätzen auf der Ebene unterschiedlicher Unterrichtsmodelle (Stürzer 2003b, Schöllermann 2003, Kasten 2010, Hannover, eingereicht).

2. Untersuchungsansatz

Die Zeugnisnoten am Ende der Klassen 5 und 6 an Haupt- und Realschulen im Regierungsbezirk Südwürttemberg¹ wurden in der Form erfasst, dass ohne Kenntnis der Namen von Schülerinnen und Schülern, von Lehrkräften oder die Klassenzugehörigkeit erkennbar war, ob das Zeugnis von einem Jungen oder Mädchen stammte, und ob die Noten von einer Lehrerin oder einem Lehrer erteilt worden waren.

Mit der Festlegung der Zeugnisnote als abhängiger Variable weichen wir allerdings von der Untersuchung von Carter (1952) insofern ab, als wir ausschließlich ein globales Fähigkeitsmaß zur Analyse heranziehen wollten, in das vielerlei Faktoren, und damit auch Faktoren der Persönlichkeit von Schülerinnen und Schülern und Lehrkräften, eingehen (vgl. Ingenkamp 1995, S. 45ff). Die von Carter (1952) verwendeten Zensuren bezogen sich ja auf einen enger umgrenzten Fähigkeitsbereich (nur Algebra). Tests und andere objektive Leistungsmessverfahren sind explizit so angelegt, dass das Messergebnis weitgehend unabhängig ist von der Person, die das Verfahren anwendet und auswertet. Bei diesen Verfahren kann folglich auch das Geschlecht der Lehrkraft praktisch keinen Einfluss auf die Leistungsbeurteilung haben. Tests werden zur möglichst objektiven Beurteilung der Leistungsfähigkeit von Schülerinnen und Schülern eingesetzt, wie z.B. in den TIMSS-, PISA- oder IGLU-Untersuchungen, bei denen es auch um die Erfassung von eventuellen Geschlechtsunterschieden auf Seiten der Schülerinnen und Schüler ging. Wie sehr Testergebnisse und Zensuren voneinander abweichen können, darauf weist bereits Carter (1952) hin. Die PISA-Ergebnisse belegten allerdings noch klarer, wie sehr sich gerade auch im Fach Mathematik Testleistungen und Zensuren voneinander unterscheiden können (s. dazu vbw 2009, S. 89 + 102) und um wie viel höher die Testleistung von Jungen sein musste, um eine den Mädchen vergleichbare Note zu erhalten. Damit war klar, dass

¹ Die Genehmigung zu dieser Untersuchung wurde vom Regierungspräsidium in Tübingen erteilt. Die Datenerhebung erfolgte durch die Studierenden, die dafür einen Leistungsnachweis erhielten.

- a) *gender*-Einflüsse auf Seiten der Schülerinnen und Schüler sich in den Zensuren weniger deutlich widerspiegeln würden als in Testergebnissen wegen der kompensatorischen Wirkung des öfter störenden Verhaltens der Jungen, bzw. des angepassteren Verhaltens der Mädchen und
- b) *gender*-Einflüsse von Seiten der Lehrkräfte vor allen hier Eingang finden könnten.

Anders als bei Testergebnissen sind Beurteilungen, also Zeugnisnoten, immer in zwei Richtungen interpretierbar. Sie sagen etwas aus über die Leistungsfähigkeit der Schülerinnen und Schüler, lassen aber auch eventuell Rückschlüsse zu auf die urteilenden Lehrkräfte und ihren eventuell vorhandenen *gender-bias*².

Insgesamt wurden die Zeugnisnoten von 3216 Kindern erhoben. Von den 1798 Hauptschulkindern waren 998 Jungen und 800 Mädchen. Bei den 1418 Realschülerinnen und -schülern handelte es sich um 699 Jungen und 719 Mädchen. Eine Überprüfung der Geschlechterverteilung bei den Haupt- und Realschülerinnen und -schülern ergab einen χ^2 -Wert von 9,03, der auf dem 1%-Niveau signifikant war. An den Hauptschulen waren die Jungen überrepräsentiert, an den Realschulen herrschte näherungsweise Gleichverteilung bei geringem Überhang an Mädchen. Diese Geschlechterverteilung steht in Übereinstimmung mit den Ergebnissen der PISA-Studie 2006 (vgl. vbw 2009, S. 96). In Bezug auf die Lehrerschaft lässt sich feststellen, dass an den Hauptschulen 58,2% aller Zensuren von Lehrern erteilt worden waren, die das Fach Mathematik unterrichteten, und 41,8% der Zensuren von Lehrerinnen. An den Realschulen lag die Geschlechterverteilung bei 51,4% männlich zu 48,6% weiblich. Eine Überprüfung der Gleichverteilung des Beurteilergeschlechts bei Haupt- und Realschulen mit dem χ^2 -Test ergab einen hochsignifikanten χ^2 -Wert_{df=1} von 15,47. Man kann also feststellen, dass an der Hauptschule deutlich mehr Zensuren von Lehrern als von Lehrerinnen erteilt worden waren, während an der Realschule die Verteilung eher ausgeglichen war. Ein Unterschied im Einsatz der Lehrkräfte an Haupt- und Realschulen könnte zudem noch eine, wenn vielleicht auch geringe Rolle gespielt haben. An Hauptschulen wird Lehrkräften nach dem Klassenlehrerprinzip vermutlich häufiger der Mathematikunterricht auch einmal fachfremd übertragen. Nach den Studienordnungen des Landes Baden-Württemberg dürften aber die meisten Lehrkräfte dann wenigstens im Grundstudium eine Art Grundausbildung im Fach Mathematik erhalten haben. Zudem können Lehrkräfte sehr wohl diesen Unterricht ablehnen, falls sie ihre eigenen Kompetenzen in diesem Bereich als zu gering einschätzen und möglicherweise haben mehr Lehrerinnen als Lehrer von diesem Recht Gebrauch gemacht.

Grundsätzlich erschiene die Anwendung einer Mehrebenenanalyse der vorliegenden Daten als wünschenswert, denn es wurden de facto ja immer die Daten von ganzen Schulklassen erfasst. Da sowohl die Lehrkräfte an den Schulen als auch das Oberschulamt verhindern wollten, dass Rückschlüsse auf einzelne Lehrkräfte oder Schulen gezogen werden könnten, wurde von beiden Seiten darauf gedrungen, dass weder Angaben zur Klassenzugehörigkeit noch zur Schulzugehörigkeit aufgenommen wurden. Das mag man bedauern, war aber in diesem Fall nicht zu ändern. Unterschiede zwischen den vielen Schulklassen bleiben also unbeachtet. Die auf die Klassenzugehörigkeit entfallende Varianz wird dadurch nicht als systematische Varianz von der Restvarianz abgezogen. Mit dem dadurch größeren Fehlerterm arbeiteten wir also in einer Varianzanalyse eher gegen uns. Für jeden Schüler war aber individuell sein Geschlecht festgehalten worden und die Tatsache, ob ein Lehrer oder eine Lehrerin die Zensur erteilt hatte.

Die Datenanalyse beinhaltet in solchen Untersuchungen zudem besondere Probleme, weil das Messniveau der untersuchten Variable „Zeugnisnoten“ genaugenommen nur Ordinalskalengüte³ hat. Bei der Anwendung der Varianzanalyse zur zufallskritischen Beurteilung von

² Der Begriff *gender-bias* bezeichnet hier Urteilsverzerrungen, deren Wirksamkeit sich die urteilende Person nicht einmal bewusst sein muss und die mit der Zugehörigkeit zu einem Geschlecht zusammenhängen.

³ Zur Skalenqualität siehe z.B. Heller & Rosemann (1974) S. 81ff.

Gruppenunterschieden verletzt man also die Voraussetzung des Intervallskalenniveaus. Wenn wir uns trotzdem für dieses Analyseverfahren entschieden, dann weil

1. in relativ vielen Untersuchungen genauso verfahren wurde mit dem Hinweis auf die Robustheit des Verfahrens und
2. bei der Untersuchung von Birkel (1978, S.240) deutlich wurde, dass im mittleren Notenbereich die Intervallunterschiede relativ gering ausfielen. Nur beim Notenschritt von der 5 zur 6 vergrößerte sich das Intervall deutlich. Man kann also vermuten, dass es sich bei der Notenskala um keine sehr gravierende Abweichung von der Intervallskalengüte handelt.

FaktorA	Hauptschule								Realschule							
FaktorB	Klasse 5				Klasse 6				Klasse 5				Klasse 6			
FaktorC	Junge		Mädchen		Junge		Mädchen		Junge		Mädchen		Junge		Mädchen	
FaktorD	L♂	L♀	L♂	L♀	L♂	L♀	L♂	L♀	L♂	L♀	L♂	L♀	L♂	L♀	L♂	L♀
N =	280	262	237	214	299	157	231	118	178	188	180	199	185	148	186	154

Tabelle 1: Varianzanalytisches Design und Zellenbesetzungen

Ein varianzanalytisches Design erschien als sinnvoll, weil damit auch die Interaktion zwischen dem Geschlecht der Schülerinnen und Schüler und dem der Lehrkraft (C*D) überprüft werden konnte. Da gleichzeitig auch die Jahrgangsstufe und die Schulart der Schüler bekannt war, ließ sich das Design zu einer 4-faktoriellen Varianzanalyse mit den Faktoren A = Schulart, B = Klassenstufe, C = Geschlecht des Lehrers und D = Geschlecht der Schülerinnen und Schüler erweitern. Abhängige Variable war die Zeugnisnote, die ordnungsgemäß immer als ganzzahliger Wert angegeben war. Wie fast immer bei solchen anfallenden Stichproben darf man nicht von gleichen Zellenbesetzungen ausgehen. Die relativ wenigen an den Hauptschulen verbliebenen Mädchen führen einmal zu einer Zellenbesetzung von 118 in der Klasse 6, während bei den Jungen mit 299 Personen die höchste Zellenbesetzung erreicht wird. Trotz dieser Unterschiede sind aber alle Zellen hinreichend häufig besetzt.

3. Hypothesen

- a) Da Lehrkräfte ihre Schülerinnen und Schüler immer im Kontext des jeweiligen Klassenverbandes zu beurteilen haben und dazu neigen dürften, die Notenskala auch auszunutzen (s. Ingenkamp 1995, S. 194ff), werden keine bedeutsamen Unterschiede zwischen den an Haupt- und Realschulen erteilten Noten erwartet.
- b) Eine generelle Veränderung der in den Klassenstufen 5 und 6 erteilten Noten wird aus dem gleichen Grund nicht erwartet.
- c) Im Gegensatz zu Carter (1952) und im Einklang mit den Testergebnissen der TIMSS- und PISA-Studien wird eine Überlegenheit der Jungen gegenüber den Mädchen im Fach Mathematik erwartet, die sich auch in besseren Zensuren niederschlägt.
- d) Dass das Lehrgeschlecht ganz allgemein die Vergabe von Zeugnisnoten im Fach Mathematik beeinflusst, wird zumindest von den Studierenden der Pädagogischen Hochschule bezweifelt, auch wenn Carter (1952) im Bereich Algebra und Klein (2004) ganz allgemein von deutlich milderem Beurteilungen durch die Lehrerinnen berichten. Unsere Erwartungen decken sich eher mit der Invarianzhypothese von Marsh et al. (2008). Unsere Untersuchungsergebnisse könnten auch einen ersten Fingerzeig geben, ob die mildere Urteilstendenz – sofern sie auftritt - eher „typisch weiblich“ oder bei deren Fehlen vielleicht doch den Fortschritten in der Professionalisierung weiblicher Lehrkräfte zuzuschreiben ist.
- e) Eine relevante Interaktion zwischen dem Geschlecht von Lehrkräften und dem der Schülerinnen und Schüler (Interaktion C*D) wird nicht erwartet, auch wenn die Ergebnisse von Carter (1952) eine solche als möglich erscheinen lassen. Auch wenn Marsh et al. (2008) die Motivation von Schülern untersuchten, so sind wir eher dazu bereit, der dort berichteten Invarianzhypothese zu folgen als der Passungshypothese.

4. Ergebnisse

Für den Faktor A = Schulart signalisiert das Ergebnis der Varianzanalyse den stärksten Effekt ($F_{1,3201} = 173,32$; $p < .001$, $Eff\% = 5,08^4$). Insgesamt gesehen erhielten die Hauptschülerinnen und -schüler mit einer Durchschnittsnote von 3,07 eine schlechtere Beurteilung als die Realschülerinnen und -schüler, bei denen der Mittelwert bei 2,65 lag. Damit ist unsere Hypothese a) abzulehnen.

Der Faktor B = Jahrgangsstufe wird erwartungsgemäß nicht signifikant ($F_{1,3201} = 1,52$, $n.s.$). Über beide Schularten gemittelt ändern sich die Beurteilungen von der 5. zur 6. Klasse unwesentlich. Damit kann unsere Hypothese b) beibehalten werden.

Bei der Beurteilung der Mittelwertsunterschiede zwischen den von Lehrerinnen und Lehrern erteilten Noten ergibt sich eine Signifikanz auf dem 5%-Niveau ($F_{1,3201} = 4,46$; $p < .05$). Im Gegensatz zu Carter (1952) urteilten aber hier die Lehrer ($M = 2,83$) einen Hauch milder als die Lehrerinnen ($M = 2,89$). Wegen der geringen Varianzklärung ($Eff\% = 0,13$) wird man aber diesen Unterschied wohl vernachlässigen und davon ausgehen können, dass hier kein wirklich bedeutender Beurteilungsunterschied in Abhängigkeit vom Lehrergeschlecht besteht. Insofern ist unsere Hypothese d) eher als bestätigt denn als widerlegt anzusehen. Hinzu kommt der Aspekt, dass der Beurteilungsunterschied zwischen Lehrerinnen und Lehrern hier in entgegen gesetzter Richtung zu den bisher berichteten Untersuchungen verläuft. Das könnte ein erster Hinweis darauf sein, dass die seinerzeit berichtete mildere Urteilstendenz der Lehrerinnen nicht unbedingt als „typisch weibliches“ Wesenselement, sondern vielleicht doch eher als Angleichung der Urteilkriterien im Zusammenhang mit der Professionalisierung zu betrachten gewesen ist.

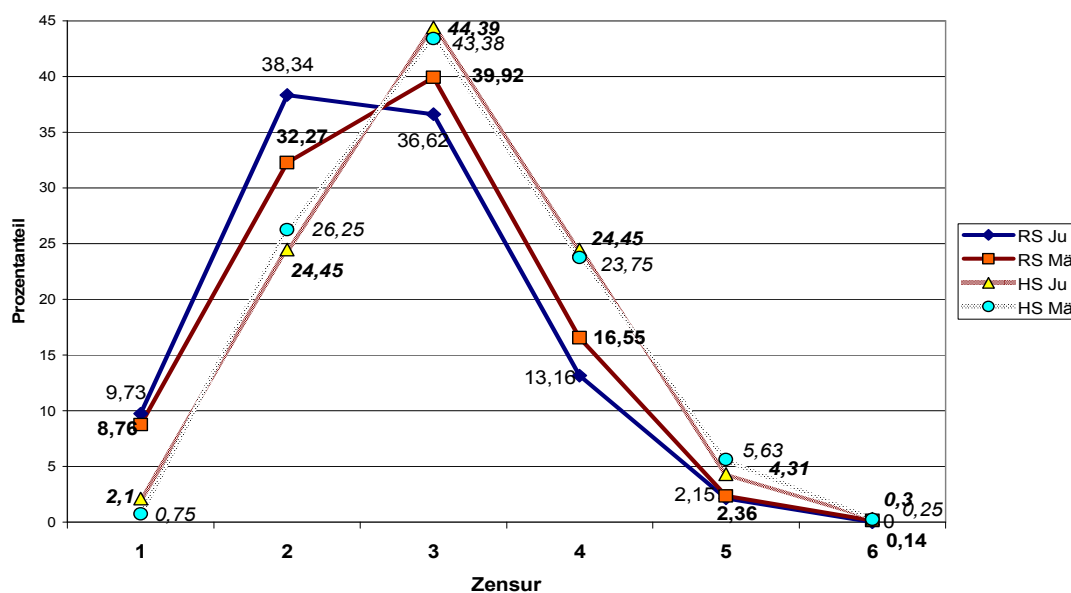


Abbildung 2: Prozentuale Notenverteilungen der Jungen und Mädchen an Haupt- und Realschulen

Mit einem Wert $F_{1,3201} = 6,54$ wird für den Faktor D = Schülgengeschlecht eine Signifikanz auf dem 1%-Niveau angezeigt. Jungen erhalten mit einem Mittelwert von 2,82 Zensuren, die signifikant über denen der Mädchen mit einem Mittel von 2,90 liegen. Unsere Hypothese c) könnte damit als prinzipiell bestätigt angesehen werden. Trotz dieser Signifikanz darf der Ge-

⁴ Eff% gibt den Anteil der geklärten Varianz für die jeweilige Varianzquelle an und entspricht in etwa dem Eta². Die Bezeichnung wurde dem Programmpaket KMSS von Kleiter (2004) entnommen.

schlechtsunterschied ebenfalls nicht überbewertet werden, denn bei einer Varianzklärung von 0,19% ist die Relevanz dieses Ergebnisses als gering einzustufen. Zur Illustration dieses Ergebnisses seien die prozentualen Notenverteilungen der Jungen und Mädchen an Haupt- und Realschulen dargestellt.

Man erkennt, dass die Notenverteilungen der Jungen und Mädchen an der Hauptschule fast identisch ausfallen und sich deutlicher nur von den Notenverteilungen der Jungen und Mädchen an den Realschulen unterscheiden (s. Schularteffekt Faktor A). Bei den Realschülerinnen und -schülern sind allerdings die Geschlechtsunterschiede etwas deutlicher ausgeprägt. Jungen empfangen an der Realschule häufiger die Noten „sehr gut“ und „gut“ und seltener „befriedigend“ und „ausreichend“ als die Mädchen. Da bei großen Stichproben bereits relativ kleine Mittelwertsunterschiede als Signifikanz erscheinen, ist in solchen Fällen eher dem Relevanzmaß zu vertrauen. Die Geschlechtsunterschiede bei den Zensuren sind also als eher wenig relevant zu betrachten.

Eine Interaktion zwischen den Faktoren C*D (Lehrergeschlecht*Schülergeschlecht) wurde nicht signifikant ($F_{1,3201} = 0,05$; *n.s.*). Der Vorsprung der Jungen betrug sowohl bei den Lehrerinnen als auch bei den Lehrern 0,08 Notenstufen und war damit wiederum recht gering. Eine deutlichere Bevorzugung der Geschlechtsgenossinnen durch die Lehrerinnen, von der Carter noch berichtet, tritt hier nicht einmal im Ansatz auf. Stattdessen wird hier der Befund von Klein (2004) bestätigt, der ebenfalls keine Interaktion zwischen dem Geschlecht von Lehrkraft und Schülerin und Schüler berichtet.

Stattdessen zeigen sich zwei andere Interaktionen als bedeutsam. Zunächst wird die Interaktion A*B (Schulart*Klassenstufe) signifikant ($F_{1,3201} = 3,73$; $p < .05$). Es kann festgestellt werden, dass der Schularteffekt auf beiden Klassenstufen in gleicher Weise – wie oben beschrieben – auftritt. Es lässt sich aber auch ein Klassenstufeneffekt an der Hauptschule beschreiben. Die Noten der Hauptschülerinnen und -schüler sacken von 3,02 in der 5. Klasse auf 3,12 in der 6. Klasse ab ($q = 3,23^3$; $Sp = 2$; $p < .05$), während sie bei den Realschülerinnen und -schülern bei einer geringfügigen Verbesserung von 2,66 auf 2,64 praktisch gleich bleiben.

Schließlich wird noch die Interaktion bei den Faktoren B(= Jahrgangsstufe) und C(= Lehrergeschlecht) signifikant ($F_{1,3201} = 6,18$; $p < .01$; $Eff\% = 0,19$). Eine Überprüfung der Mittelwertsunterschiede ergab einerseits einen Jahrgangsstufeneffekt bei den männlichen Lehrern ($q = 3,96$; $Sp = 3$; $p < .05$), die in der 6. Klasse ($M = 2,89$) strengere Zensuren erteilten als in der 5. Klasse ($M = 2,77$), und andererseits einen Effekt des Lehrergeschlechts in der 5. Klasse ($q = 4,47$; $Sp = 4$; $p < .01$). In Klasse 5 gaben die Lehrer ($M = 2,77$) deutlich mildere Zensuren als die Lehrerinnen ($M = 2,91$). Daraus lassen sich aber wohl auch kaum allgemeine Urteilstendenzen der Lehrkräfte ablesen, zumal der Mittelwertsunterschied in Klasse 6 tendenziell anders herum auftrat. Der Einfluss des Haupteffekts Lehrergeschlecht (Faktor C) wird also vor allem durch die Lehrkräfte verursacht, die in der 5. Jahrgangsstufe unterrichteten, während er bereits in der 6. Jahrgangsstufe wieder verschwunden zu sein scheint.

5. Diskussion

Man wird sicher *nicht* davon ausgehen dürfen, dass die Problematik der Zensurenggebung heute als gelöst angesehen werden darf (s. Birkel 2003, 2005, 2009, Tent & Birkel 2010), aber einige Veränderungen im Bereich der Schule scheinen sich doch auf die Zensurenggebung auszuwirken. Eine solche Veränderung ist vermutlich die Tatsache, dass sich seit 1952 die Anzahl der in Schule und Unterricht tätigen Lehrerinnen deutlich im Vergleich zu den männlichen Kollegen verändert hat. Der sich immer weiter erhöhende Anteil von Lehrerinnen an den Kollegien dürfte zu einem – vermutlich meist impliziten – Aushandlungsprozess bezüglich der Kriterien für die

⁵ Die Prüfstatistik q resultiert aus dem Spannweitentest nach Newman-Keuls, wie er bei Heller & Rosemann (1974, S.205ff) beschrieben ist.

Beurteilung von Schülerleistungen beigetragen haben, der zu ihrer stärkeren Angleichung zwischen den Geschlechtern führte. Man kann also vermuten, dass die Veränderung der Geschlechterverhältnisse in den Kollegien auch zu Auswirkungen auf den Professionalisierungsprozess bei Lehrerinnen und Lehrern führte. So darf man wohl davon ausgehen, dass noch in den 50er-Jahren ein deutliches Übergewicht männlicher Lehrer in den Lehrerkollegien vorherrschte. Frauen mussten sich vermutlich in ihrem Beruf als Lehrerinnen noch etablieren. Eine Aussage, dass die Tendenz, mildere Zensuren zu erteilen auf diesem Hintergrund zu verstehen sei, verschaffe doch die Vergabe besserer Zensuren den Lehrerinnen eventuell mehr Anerkennung und Ansehen bei den Schülern, Eltern und Schulträgern, denn man könne die besseren Zensuren ja auch interpretieren als Beleg für den erfolgreicherem Unterricht, erscheint aber im Moment noch als ebenso spekulativ wie die Interpretation, dass die etablierten männlichen Kollegen geglaubt haben könnten, sich strengere Maßstäbe bei der Beurteilung der Mathematikleistungen erlauben zu können, weil ihre Professionalität nicht kritisch hinterfragt wurde. Der Veränderungsprozess der Beurteilungen von Lehrerinnen und Lehrern hin zu einer Angleichung spräche also eher für eine Etablierung der Lehrerinnen als anerkannte pädagogische Fachkräfte. Nur die Bestätigung der seinerzeit von Klein (2004) berichteten Beurteilungsunterschiede hätte für eine dem Stereotyp „typisch weibliche Lehrkraft“ entsprechende Beurteilungstendenz gesprochen. Es darf somit bezweifelt werden, dass die mildere Urteilstendenz immer noch konstitutiv zum Stereotyp der Lehrerin gehört. Unsere Ergebnisse stützen eher die Invarianzhypothese, die Marsh et al. (2008) im Bereich der Motivation der Schülerinnen und Schüler durch gleich- oder gegengeschlechtliche Lehrkräfte gefunden haben.

Die unterschiedliche Beurteilung von Schülerleistungen durch Lehrerinnen und Lehrer, wie sie seinerzeit von Carter (1952) und relativ neu von Klein (2004) dokumentiert wird, wurde noch vor 20 Jahren von Flaake (1989) im Hinblick auf die psychoanalytisch interpretierte Sozialisation der Geschlechter diskutiert. Bei ihr scheint die mildere Notengebung der Lehrerinnen dafür zu sprechen, dass sie ihre Rolle in der Schule als eher empathisch und damit näher am Kind verstehen. Lehrerinnen könnten sich eher in die psychische Verfassung von Schülern einfühlen, die schlechte Noten erhalten. Ob diese generelle Betrachtung von Lehrerinnen allerdings gerechtfertigt werden kann, ist bisher nicht überprüft. So könnte man sich fragen, ob das von den Lehrerinnen gewählte Unterrichtsfach nicht auch Auswirkungen auf das Urteilsverhalten haben könnte. Die sogar etwas strengere Urteilstendenz der Lehrerinnen im Fach Mathematik, wie sie hier gefunden wurde, scheint zunächst einmal gegen die Interpretation von Flaake (1989) zu sprechen. Zu überprüfen wäre nun, ob die hier gefundenen Verhältnisse bei der Schülerbeurteilung im Fach Mathematik auf alle Unterrichtsfächer generalisiert werden dürfen. Man muss vielleicht doch fragen, welche junge Frau entscheidet sich zu einem Studium des Faches Mathematik. Entgegen dem allgemeinen Trend haben die männlichen Studierenden in diesem Studienfach ein deutliches Übergewicht. Sind es die gleichen Frauen, die sich auch für andere Studienfächer hätten entscheiden können? Könnte es sein, dass Frauen mit einer besonderen Persönlichkeitsstruktur (einer stärker der männlichen angeglichenen) sich für ein solches Studium entscheiden, und darum auch in ihrem Urteilsverhalten keine wesentlichen Unterschiede zu den Lehrern zeigen? Solche Fragen könnten Anlass zu weiteren Untersuchungen bieten. Es könnte also durchaus sein, dass *gender*-Differenzen bei der Schülerbeurteilung je nach Unterrichtsfach unterschiedlich ausfallen. Lehrerinnen könnten ihre Rolle im Fach Mathematik anders definieren als in anderen Fächern wie z.B. Deutsch oder Musik. Ergebnisse einer Analyse von Deutschnoten (Birkel 2011) scheinen dafür zu sprechen.

Möglicherweise trägt aber auch die meist unterstellte relative Objektivität der Leistungsmessungen im Fach Mathematik⁶ dazu bei, dass die Beurteilungen durch Lehrerinnen und Lehrer eher ähnlich ausfallen. Zwar gehen in die Zeugnisnoten viele Faktoren ein, aber die schriftlichen Leistungsüberprüfungen spielen bei der Festlegung der Note doch eine prominente Rolle. Die

⁶ Die ja so objektiv wie meist erwartet gar nicht ist. Siehe dazu Birkel (2005).

Richtigkeit oder Falschheit von Aufgabenlösungen ist vergleichsweise übereinstimmend feststellbar und führt zumindest im Klassenverband zu einer oft auch für die Schüler transparenten Zuordnung von Leistungsergebnissen zu Zensuren. Warum also sollten hier bedeutsame Geschlechtsunterschiede auftreten? Bei aller Problematik der Notengebung auch im Fach Mathematik, wie sie bei Birkel (2005) dargestellt wurde, könnte die Struktur der Leistungsüberprüfungen doch dazu beitragen, dass die Beurteilungsunterschiede zwischen den Lehrerinnen und Lehrern gering ausfallen oder sich sogar einmal tendenziell ins Gegenteil verkehren.

Die immer stärker werdende Femininisierung des Lehrerberufs hat sicher zu einem veränderten Verständnis der Lehrerinnenrolle geführt (siehe auch Horstkemper 2000, Essen & Rogers 2006). Lehrerinnen als gleichberechtigte Mitglieder im Kollegium werden heute als selbstverständlich und in ihrer Professionalität als den männlichen Kollegen nicht nur gleichwertig, sondern vor allem in der Anwendung schülerorientierter Unterrichtsformen ihnen sogar als überlegen betrachtet. Händle (1989) betrachtet sie gar als „*Speerspitze des pädagogischen Fortschritts*“. Bei der Untersuchung hier urteilten die Lehrerinnen sogar signifikant (wenn auch nicht besonders stark!) strenger als die Lehrer. Auch der Effekt des Faktors Lehrergeschlecht bei den Fünftklässlern in der Interaktion mit dem Faktor Jahrgangsstufe weist die Lehrerinnen als etwas strengere Beurteiler aus. Insgesamt scheint hier das Lehrergeschlecht keine wirklich relevante Rolle im Hinblick auf die Notenerteilung gespielt zu haben. Klein (2004) stellte die Vermutung auf, dass die Lehrer stärker Verhalten und Leistung in der Urteilsbildung miteinander vermischen, während Lehrerinnen diese Aspekte bewusster trennen. Ob diese Vermutungen auch auf die hier untersuchten Lehrerinnen und Lehrer zutrifft, ist eher zu bezweifeln. Dafür könnten auch nationale Unterschiede verantwortlich sein. Weitere Untersuchungen in Deutschland müssten sich auf die *genderspezifischen* Unterschiede bei der Urteilsfindung im Notengebungsprozess konzentrieren.

Die von Carter (1952) berichtete Interaktion zwischen dem Geschlecht der Lehrkräfte und dem der Schülerinnen und Schüler scheint zumindest in Bezug zu den erteilten Zeugniszensuren keine Gültigkeit zu besitzen. Unsere Ergebnisse zeigen eher Übereinstimmung mit den Befunden von Klein (2004) und Marsh et al. (2008). Wenn man versuchen will zu verstehen, warum seinerzeit bei Carter die Lehrerinnen die Mädchen so besonders deutlich besser beurteilten in einem Fach wie der Algebra (vielleicht urteilten ja auch die Lehrer zu streng?), dann muss man sich vielleicht in Erinnerung rufen, dass damals in Deutschland die Diskussion um die Benachteiligung von Mädchen in der Schule begann. Wollte man es den Lehrerinnen heute noch verübeln, wenn sie damals der Benachteiligung der Mädchen möglicherweise gegenzusteuern versuchten? Fast 60 Jahre lang ging es doch darum, eventuelle Benachteiligungen von Mädchen im Schulsystem auszugleichen. Mädchen sind inzwischen die erfolgreicheren Schülerinnen geworden. Es gibt nur wenige Bereiche, in denen die Jungen erfolgreicher abschneiden, Mathematik gehört in Deutschland dazu. Heute gibt es schon eher wieder eine Gegenbewegung beim Abbau von deren Benachteiligungen im Schulsystem, wenn von der „*Metamorphose der Arbeitertochter zum Migrantensohn*“ (Geißler 2008) gesprochen wird. Hannover (eingereicht) fragt sich gar, ob nicht die Überrepräsentanz von Frauen im Bildungssystem die Jungen zu Bildungsverlierern macht. Die *gender*-Problematik auf Seiten der Schülerinnen und Schüler ist inzwischen recht gut, auf Seiten der Lehrkräfte aber fast gar nicht untersucht. Ob und wie sich das Geschlecht der Lehrkräfte und deren Interaktion mit dem Geschlecht der Schülerinnen und Schüler auf Beurteilungsprozesse auswirken, bedarf noch intensiver Erforschung.

Die Signifikanz des Faktors C (= Schülergeschlecht) deutet darauf hin, dass das Fach Mathematik erwartungsgemäß wohl eher eine Jungendomäne ist. Der Notenunterschied ist aber noch recht gering und wenig relevant. In den Klassen 5 und 6 schreitet aber offensichtlich die Ausdifferenzierung geschlechtsspezifischer Fachinteressen erst noch voran, von der bereits Tiedemann & Faber (1994) als beginnend in der Grundschule berichten, und die im weiteren Schulverlauf eher deutlicher wird (vgl. Baumert 1997, 2000). Die etwas besseren Zensuren der Jungen im Fach Mathematik hier stehen dabei in Einklang mit ihren besseren Testleistungen bei den

internationalen Vergleichsuntersuchungen (vgl. Deutsches PISA-Konsortium 2001, 2004, 2007, zusammenfassend vbw 2009), müssen aber bewertet werden auf dem Hintergrund einer noch stärkeren Überlegenheit der Mädchen in den sprachlichen Fächern. Die geringe Überlegenheit der Jungen bei den Zensuren könnte einerseits mit ihrem weniger angepassten Verhalten in schulischen Leistungssituationen zusammenhängen, andererseits aber auch mit der Qualität der Mathematikaufgaben in den Klassen 5 und 6. In diesen Jahrgängen sind vielleicht noch nicht so viel abstrakte mathematische Fähigkeiten erforderlich wie in späteren Jahrgangsstufen, in denen die männliche Dominanz stärker hervortritt. Die genannten Fähigkeiten werden von Mädchen offenbar als weniger erstrebenswert und nicht mit dem *gender*-Stereotyp kompatibel (Steffens et al. 2010) erlebt und deshalb im weiteren Verlauf der Schulzeit nicht in dem Maße ausgeprägt wie bei den Jungen (Nyssen 2004, Kampshoff 2007). Implizite *gender*-Stereotype bezüglich der Bedeutung von Mathematik beeinflussen das Selbstkonzept und die schulischen Leistungen von Mädchen in Mathematik bereits ab einem Alter von neun Jahren und gewinnen im Verlauf der Entwicklung weiter an Bedeutung (Steffens et al. 2010). Solche Stereotype können u.U. von den Mädchen sogar als bedrohlich und damit als leistungshemmend erlebt werden (zum „*stereotype threat*“ siehe Good 2001).

Wenn der Geschlechtsunterschied im Fach Mathematik hier nicht so markant hervortritt wie erwartet, dann ist eben auch zu bedenken, dass in den internationalen Vergleichsstudien Tests zur Abbildung mathematischer Fähigkeiten verwendet wurden. Auf die Diskrepanzen zwischen Testleistungen und Zensuren im Fach Mathematik machen u.a. Schreiner, Breit & Haider (2008) anhand der PISA-Ergebnisse für Österreich aufmerksam, wenn sie feststellen dass „*die Burschen ... am Ende der Pflichtschulzeit keine besseren Noten erhalten als ihre Kolleginnen, obwohl sie 1 oder 2 Jahre später bei PISA besser abschneiden als ihre Mitschülerinnen*“ (S.216). Zu überlegen wäre also, warum die vor allem in Deutschland in Leistungstests immer wieder sichtbaren Vorteile der Jungen⁷ (z.B. LAU → Lehmann & Peek 1997, PISA → Dt. PISA-Konsortium 2007) sich in den Zensuren, die hier vergeben wurden, nicht so deutlich niederschlagen. Testleistungen müssen nicht den Forderungen der Schuladministration genügen, die Leistungen im Verlauf eines Schuljahres in all ihren Facetten abzubilden, von der mündlichen Beteiligung über Kurztests bis hin zu aufwändigen Klassenarbeiten. Genau das und einige weitere Funktionen soll aber ausdrücklich die Schulnote erfüllen. Von daher darf es also nicht verwundern, dass die bessere Anpassung der Mädchen an schulische Verhaltensnormen (s. Spinath et al. 2010) zu gewissen kompensatorischen Effekten bei der Mathematiknote geführt haben könnte. Dazu schreibt Budde (2009, S. 35), dass es sich negativ auf die Benotung auswirke, dass Jungen im Unterricht häufiger unangenehm auffallen, während Mädchen sich viel schulangepasster verhalten⁸. Er belegt das mit dem Vergleich von Kompetenzwerten und Schulnoten von Fünftklässlern aus der QuaSUM-Studie (Lehmann et al. 1999b). Bei gleichen Kompetenzwerten bekommen Mädchen bessere Zensuren als Jungen, und das bestätigt Budde (2009) auch mit eigenen Untersuchungsergebnissen. Wenn also die Zensurenunterschiede in dieser Untersuchung durch solche Urteilstendenzen der Lehrkräfte überlagert sind, dann deuten die zumindest etwas besseren Zensuren der Jungen vielleicht doch auf noch deutlichere Kompetenzunterschiede im Vergleich zu Mädchen hin. Auch die höhere Korrelation zwischen Testleistung und Zensur (Budde 2009) bei den Jungen ($d = .57$) im Vergleich zu den Mädchen ($d = .47$) ließe sich in der Richtung interpretieren, dass in die Zensuren der Mädchen mehr leistungsfremde (Mädchen bevorzugende?) Faktoren eingehen als bei Jungen. Budde (2009, S. 36) zieht daraus den Schluss, „dass es einen Bias in der Bewertung von Jungen und Mädchen zu geben scheint, der zu Gunsten

⁷ Zum Zusammenhang zwischen Kompetenzunterschieden und Schülergeschlecht sei auf Guiso et al. (2008) verwiesen, der verdeutlicht, wie sehr diese Differenzen korrelieren mit Maßen für die weibliche Emanzipation in der Gesellschaft. In Deutschland korreliert der vergleichsweise niedrige Emanzipationswert mit deutlichen Geschlechtsunterschieden im Bereich der mathematisch-naturwissenschaftlichen Kompetenzen.

⁸ Schulunangepasstes Verhalten führt bei Jungen und bei Mädchen zu einer negativeren Beurteilung der Matheleistung. Dieses Verhalten kommt nur bei Jungen viel häufiger vor.

der Mädchen ausfällt. Bei Jungen decken sich Kompetenzwerte aus Tests und Zeugnisnoten stärker als bei Mädchen, d.h. Jungen werden ‚objektiver‘ beurteilt“.

Vielleicht sind aber hier bereits erste Erfolge zu erkennen, die darauf abzielen, den Mathematikunterricht *gender*-sensibel so zu gestalten (Budde, Scholand & Faulstich-Wieland 2008), dass ein Erfolg der Mädchen wahrscheinlicher wird. Guiso et al. (2008) weisen zudem darauf hin, dass in Ländern mit einer relativ ausgeprägten Chancengleichheit für beide Geschlechter die Geschlechtsunterschiede bei den Mathematikleistungen sich deutlich verringern. Generell korrelierte der Index für weibliche Emanzipation (GGI) in den Teilnehmerländern der PISA-Studien zu -0.57 mit der Stärke der männlichen Dominanz im Fach Mathematik. Für Deutschland bedeutet die immer noch ausgeprägte Überlegenheit der Jungen in den MINT-Fächern⁹ (s. auch Quaiser-Pohl & Endepohls-Ulpe 2010), dass im beruflichen Leben selbst hierin begabte Mädchen Karriereschwierigkeiten haben. Auch wenn inzwischen fast gleich viele Mädchen und Jungen einen Hochschulabschluss in diesen Fächern erreichen, so treten doch immer deutlicher werdende Geschlechtsdisparitäten in den weiteren Stadien von akademischen Laufbahnen auf (Stöger & Sontag 2009). Versuche, die unterschiedliche Entwicklung von Jungen und Mädchen in den MINT-Fächern zu erklären, stehen noch relativ am Anfang. Eine Auflistung möglicher Faktoren findet sich bei Hannover (eingereicht), Kasten (2010), Stöger & Sontag (2009) oder vbw (2009).

Ein interessanter Zugang zur *gender*-Problematik auf Seiten der Schüler ist der Versuch, Persönlichkeitsmerkmale zu identifizieren, die das domänen-spezifische Abschneiden von Jungen und Mädchen erklären können. Freudenthaler, Spinath & Neubauer (2008) sowie Spinath, Freudenthaler & Neubauer (2010) widmen sich diesen Fragen und versuchen so Anhaltspunkte zu finden zur Erklärung von Schülerleistungen. Sie stellen heraus, dass für beide Geschlechter bei den Schülern sowohl bei der allgemeinen Schulleistung (einer Art GPA) als auch bei verschiedenen Unterrichtsfächern die Intelligenz der beste Prädiktor für die schulischen Leistungen war. Für die Vorhersage des allgemeinen Schulerfolgs (GPA) erwies sich zudem das Leistungselbstkonzept bei beiden Geschlechtern als hoch bedeutsam, während die schulbezogene intrinsische Motivation, die Schulangst und eine Leistungs-Vermeidungstendenz die Schulleistung von Jungen vorhersagbar machte und Arbeitsvermeidungsverhalten die der Mädchen (Freudenthaler et al. 2008). Betrachtete man bei den einzelnen Schulfächern (Spinath et al. 2010) die Mathematik, so erwiesen sich sowohl die „Big Five“-Persönlichkeitsfaktoren als auch die mathespezifische Schulangst, das Fähigkeitsselbstkonzept und das Interesse am Fach sowie Gewissenhaftigkeit und Neurotizismus als wichtige Prädiktoren für beide Geschlechter. Bei den sprachlichen Fächern Deutsch und Englisch dagegen war zu erkennen, dass der Faktor Extraversion mit dem Schülergeschlecht interagierte. Hohe Extraversionswerte gingen bei den Mädchen mit guten Noten einher, während es bei den Jungen genau umgekehrt war. Extravertiertes Verhalten bei ihnen drückte sich wohl eher als Verletzung von schulischen Verhaltensnormen aus.

Ein Erklärungsversuch wäre noch zu starten bei der Signifikanz des Faktors A(= Schultart). Welche Gründe könnten dafür herangezogen werden, um die besseren Zensuren der Realschülerinnen und –schüler gegenüber denen der Hauptschülerinnen und –schüler zu verstehen? Eine Vermutung könnte in die Richtung gehen, dass das Bewusstsein der Hauptschullehrkräfte, es mit den schwächsten Schülern nach der Klasse 4 zu tun zu haben, zu den negativeren Beurteilungen beiträgt, während die Realschullehrkräfte im Bewusstsein urteilen, es mit einer eher positiven Auslese zu tun zu haben.

Ein zweiter Gesichtspunkt könnte in der Zusammensetzung der Klassen an Haupt- und Realschulen begründet liegen. In den Hauptschulklassen dieser Untersuchung befanden sich 55,5% Jungen und nur 44,5% Mädchen, während an den Realschulen der Jungenanteil bei nur 49,3%, der Mädchenanteil aber bei 50,7% lag. Da Jungen in der Schule allgemein weniger gut angepasst

⁹ MINT steht für **M**athematik, **I**nformatik, **N**aturwissenschaften und **T**echnik.

reagieren, weniger unterrichtliche Aktivitäten unterstützen und eher Disziplinprobleme bereiten als Mädchen (vgl. z.B. Nyssen 2004, Kampshoff 2007, Budde 2009, Spinath et al. 2010), werden ihre Leistungen schlechter beurteilt als die der Mädchen (zur Disziplinierungsfunktion der Noten s. Ingenkamp 1995, Beck 1974). Der höhere Anteil an Jungen in der Hauptschule und die Selektion der angepassteren Jungen für die Realschule könnte also dazu beigetragen haben, dass die Hauptschülerinnen und -schüler insgesamt negativer beurteilt werden als die Realschülerinnen und -schüler. Dieser Trend setzte sich sogar von der 5. zur 6. Klasse noch fort. Besteht in der 5. Klasse eine Notendifferenz von 0,36 Notenstufen ($q = 11,90$; $Sp = 2$; $p < .001$) so steigt sie in der 6. Klasse auf 0,48 Notenstufen an ($q = 14,15$; $Sp = 4$; $p < .001$). In der Tat sinkt der F-Wert des Haupteffekts in der Varianzanalyse ($F = 173,32$) bei Auspartialisierung des Geschlechts etwas ab auf $F = 170,11$ in einer Kovarianzanalyse. Damit lässt sich aber festhalten, dass die Geschlechtszugehörigkeit der Schülerinnen und Schüler die Signifikanz des Faktors Schulart nicht wesentlich beeinflusst hat.

Eigentlich hätte man auch erwarten können, dass sich die Schülerinnen und Schüler in ihrer Leistung an der Hauptschule neu orientieren, zumal die Konkurrenz durch die Leistungsstärkeren hier fortfällt, die an Realschule und Gymnasium überwechselten. Das Herausbilden einer neuen Leistungsspitze scheint also ein Prozess zu sein, der einfach länger dauert und nicht bereits in Klasse 6 auf die Beurteilung der Leistungen im Zeugnis durchschlägt. Hinzu kommt die Überlegung, dass die öffentliche Diskussion um das Negativ-Image der Hauptschule als „Restschule“ nicht nur bei den Lehrkräften, sondern auch bei den Schülerinnen und Schülern zu einem Motivationsverlust führt. Warum sollte die Lehrkraft Hauptschülerinnen und -schüler durch positivere Zensuren zu vermehrter Anstrengung motivieren, wenn man es mit einer „Negativauswahl“ zu tun hat? Warum sollte man sich als Schülerin oder Schüler mehr anstrengen, wenn man eben doch zu dieser „Negativauswahl“ gehört? Die Möglichkeit, einen mittleren Bildungsabschluss an der Hauptschule zu erreichen, was in Baden-Württemberg vergleichsweise häufig vorkommt, tritt vielleicht erst später ins Bewusstsein von Schülerinnen, Schülern und Lehrkräften.

6. Literatur

- Baumert, J. et al. (1997). *TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde*. Opladen: Leske & Budrich.
- Baumert, J. et al. (2000). *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Band 1: Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit*. Opladen: Leske & Budrich.
- Beck, J. (1974). *lernen in der klassenschule*. Reinbeck: Rowohlt.
- Birkel, P. (1978). *Mündliche Prüfungen. Zur Objektivität und Validität der Leistungsbeurteilung*. Bochum: Kamp.
- Birkel, P. (2003). Aufsatzbeurteilung – ein altes Problem neu untersucht. *Didaktik Deutsch*, 9, H. 15, S. 46-53.
- Birkel, P. (2005). Beurteilungsübereinstimmung bei Mathematikarbeiten? *Journal für Mathematik-Didaktik*, 26, H.1, S. 28-51.
- Birkel, P. (2009). Rechtschreibleistung im Diktat – eine objektiv beurteilbare Leistung? *Didaktik Deutsch*, 15, H. 27, S. 5-32.
- Birkel, P. (2011). Gender-spezifische Effekte bei Deutschnoten im Zeugnis? Manuskript: PH Weingarten.
- Brandon, P.R., Newton, B.J. & Hammond, O.W. (1985): *The superiority of girls over boys in mathematics achievement in Hawaii*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.
- Budde, J.; Scholand, B. & Faulstich-Wieland, H. (2008). *Geschlechtergerechtigkeit in der Schule. Eine Studie zu Chancen, Blockaden und Perspektiven einer gender-sensiblen Schulkultur*. Weinheim, München: Juventa.
- Budde, J. (2009). *Mathematikunterricht und Geschlecht. Empirische Ergebnisse und pädagogische Ansätze*. Bildungsforschung Band 30, herausgeg. vom Bundesministerium für Bildung und Forschung. Bonn, Berlin.

Peter Birkel: Geschlechtsspezifische Effekte bei der Zensurenggebung im Fach Mathematik? 14
Version 01.12.2011

- Callahan, L.G. & Clements, D.H. (1984). Sex differences in rote-counting ability on entry to first grade: Some observations. *Journal for Research in Mathematics Education*, 15, S. 378-382.
- Carter, R.S. (1952). How invalid are marks assigned by teachers? *The Journal of Educational Psychology*, 43, No. 4, S. 218-228. In deutscher Übersetzung und gekürzt erstmals abgedruckt unter dem Titel „Wie gültig sind durch Lehrer erteilte Zensuren?“ In: Ingenkamp, K. (Hrsg.) (1971¹, 1995⁹). *Die Fragwürdigkeit der Zensurenggebung*. S. 148-158. Weinheim: Beltz.
- Crotti, C. (2006). Ist der Bildungserfolg bzw. Misserfolg eine Geschlechterfrage? *Zeitschrift für Pädagogik*, 52, H. 3, S. 363-374.
- Day, L.C. (1938). Boys and girls and honour ranks. *School Review*, 46, S. 288-299.
- Deutsches PISA-Konsortium (Hrsg.). (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske & Budrich.
- Deutsches PISA-Konsortium (Hrsg.). (2004). *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs*. Münster: Waxmann.
- Deutsches PISA-Konsortium (Hrsg.). (2007). *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie*. Münster: Waxmann.
- Dossey, J.A., Mullis, I.V.S., Lindquist, M.M. & Chambers, D.L. (1988). *The mathematics report card: Are we measuring up? Trends and achievement based on the 1986 National Assessment*. Princeton: Educational Testing Service.
- Douglass, H.R. (1937). Relation of high-school marks to sex in four Minnesota senior high schools. *School Review*, 45, S. 282-288.
- Edminston, R.W. (1943). Do teachers show partiality toward boys or girls? *Peabody Journal of Education*, 20, S. 234-238.
- Essen, M. van & Rogers, R. (2006). Zur Geschichte der Lehrkräfte: Historiographische Herausforderungen und internationale Perspektiven. *Zeitschrift für Pädagogik*, 52, H. 3, S. 319-337.
- Flaake, K. (1989). *Berufliche Orientierung von Lehrerinnen und Lehrern. Eine empirische Untersuchung*. München: Campus.
- Garner, C.E. (1935). Survey of Teachers' Marks. *School and Community*, 21, S. 116-117.
- Geißler, R. (2008). Die Metamorphose der Arbeitertochter zum Migrantensohn. Zum Wandel der Chancenstruktur im Bildungssystem nach Schicht, Geschlecht, Ethnie und deren Verknüpfungen. In: P.A. Berger & H. Kahlert (Hrsg.). *Institutionalisierte Ungleichheiten. Wie das Bildungssystem Chancen blockiert*. S. 71-100. Juventa: Weinheim.
- Glaser E.; Klika D. & Prengel A. (Hrsg.). (2004). *Handbuch Gender und Erziehungswissenschaft*. Bad Heilbrunn: Klinkhardt.
- Good, C. (2001). *Stereotype threat and its relation to theories of elementary girls' mathematics achievement and task choices*. Austin/Texas.
- Guiso, L.; Monte, F.; Spienza, P. & Zingales, L. (2008). Culture, Gender and Math. *Science*, 320, 5880, S. 1164-1165.
- Händle, C. (1989). Soziale Handlungsprioritäten im Beruf. Befunde aus Fallstudien von Lehrerinnen in der Berufseingangsphase. In: H. Faulstich-Wieland (Hrsg.). *Weibliche Identität*. Bielefeld.
- Hannover, B. (eingereicht). Ist die Überrepräsentanz von Frauen im Bildungssystem für den geringeren Bildungserfolg von Jungen verantwortlich? Eine Diskussion der empirischen Evidenz. In: E.H. Witte & J. Doll (Hrsg.). *Sozialpsychologie, Sozialisation und Schule*. Lengerich: Pabst.
- Hawn, H.C., Ellet, C.D. & Des Jardines, L. (1991). *Differences in mathematics achievement between males and females in grades 1-3*. Paper presented at the annual meeting of the Eastern Educational Research Association, Philadelphia.
- Heller, K. & Rosemann, B. (1974). *Planung und Auswertung empirischer Untersuchungen*. Stuttgart: Klett.
- Horstkemper, M. (2000). Geschlecht und Professionalität. Lehrerinnen und Lehrer – Über die Bedeutung der Geschlechtsdifferenz. In: J. Bastian, W. Helsper, S. Reh & C. Schelle (Hrsg.). *Professionalisierung im Lehrerberuf*. S. 87-105. Opladen: Leske & Budrich.
- Ingenkamp, K. (Hrsg.) (1971¹, 1995⁹). *Die Fragwürdigkeit der Zensurenggebung*. Beltz: Weinheim.

Peter Birkel: Geschlechtsspezifische Effekte bei der Zensurenggebung im Fach Mathematik? 15
Version 01.12.2011

- Jahnke-Klein, S. (2006). Mathematik, Informatik, Naturwissenschaften und Technik – (noch immer) nichts für Mädchen? In: S. Jösting & M. Seemann (Hrsg.). *Gender und Schule. Geschlechterverhältnisse in Theorie und schulischer Praxis*. S.97-120. Oldenburg: BIS-Verlag.
- Kampshoff, M. (2007). *Geschlechterdifferenz und Schulleistung*. Verlag für Sozialwissenschaften: Wiesbaden.
- Kasten, H. (2010). Geschlechtsunterschiede. In: D.H. Rost (Hrsg.). *Handwörterbuch Pädagogische Psychologie*. S. 234-241. Weinheim: Beltz.
- Klein, J. (2004). Who is most responsible for gender differences in scholastic achievement: pupils or teachers? *Educational Research*, 49, Heft 2, S. 183-193.
- Kleiter, E. (2004). *KMSS-8. Kleiter-Microcomputer-Statistik-System*. Kiel
- Lehmann, R.H. & Peek, R. (1997). *Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an Hamburger Schulen. Bericht über die Untersuchung im September 1996 (unveröffentlichter Forschungsbericht)*. Hamburg.
- Lehmann, R.H.; Gänsfuß, R. & Peek, R. (1999a): *Aspekte der Lernausgangslage und der Lernentwicklung von Schülerinnen und Schülern an Hamburger Schulen – Klassenstufe 7. Bericht über die Untersuchung im September 1998 (unveröffentlichter Forschungsbericht)*. Hamburg.
- Lehmann, R.H.; Peek, R.; Gänsfuß, R.; Lutkat, S.; Mücke, St. & Barth, I. (1999b). Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik. Berlin. Abrufbar unter: www2.hu-berlin.de/empir_bf/QuaSUM.pdf
- Lewis, J.C. & Hoover, H.D. (1986). *Differential prediction of academic achievement in elementary and junior high school by sex*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- Marsh, H.W.; Martin, A.J. & Cheng, J.H.S. (2008). A Multilevel Perspective on Gender in Classroom Motivation and Climate: Potential Benefits of Male Teachers for Boys? *Journal of Educational Psychology*, 100, S. 78-95.
- McKay, L.R. (1979). Sex differences in cognitive styles and mathematics achievement in fourth and eighth graders. *Dissertation Abstracts International*, 39, 7301A.
- Potter M.C. & Lewy, E. (1968). Spatial enumeration without counting. *Child Development*, 39, S. 265-272.
- Quaiser-Pohl, C. & Endenpohls-Ulpe, M. (Hrsg.). (2010). *Bildungsprozesse im MINT-Bereich*. Münster: Waxmann.
- Nyssen, E. (2004). Gender in den Sekundarstufen. In: E. Glaser, D. Klika & A. Prengel (Hrsg.). *Handbuch Gender und Erziehungswissenschaft*. S. 391-401. Bad Heilbrunn: Klinkhardt.
- Shinnerer, M.C. (1944). Failure ratio: Two boys to one girl. *Clearing House*, 18, S. 264-270.
- Ratzka, N. (2003). *Mathematische Fähigkeiten und Fertigkeiten am Ende der Grundschulzeit. Empirische Studien im Anschluss an TIMSS*. Hildesheim: Franzbecker.
- Schöllermann, A. (2003). *Leistungsdifferenzen und Interessenunterschiede (zwischen Mädchen und Jungen) im Mathematikunterricht ab der 7. Klasse – Ursachen und Lösungsvorschläge*. http://www.erzwiss.uni-hamburg.de/degendering_science/ (11.11.2009)
- Schreiner, C.; Breit, S. & Haider, G. (2008). Zur Validität der Mathematiknoten. Ein Vergleich von Lehrerbeurteilung und Leistungsmessung bei PISA. In: F. Hofmann, C. Schreiner & J. Thonhauser: *Qualitative und quantitative Aspekte. Zu ihrer Komplementarität in der erziehungswissenschaftlichen Forschung*. S. 211-223. Münster: Waxmann.
- Shipman, V.C. (1972). *Disadvantaged Children and their first school experiences*. Princeton: Educational Testing Service, Head Start Longitudinal Study Report PR-72-18.
- Siegel, L.S. (1968). The development of the ability to process information. *Journal of Experimental Child Psychology*, 6, S. 368-383.
- Steffens, M.C.; Jelenec, P. & Noack, P. (2010). On the leaky math pipeline: Comparing implicit math-gender stereotypes and math withdrawal in female and male children and adolescents. *Journal of Educational Psychology*, 102, Sept 20, no pagination.
- Stöger, H. & Sontag, C. (2009). Geschlechtsdisparitäten im Bildungsbereich: Die Situation hochleistender und hochbegabter Mädchen und Frauen. *News & Science*, 3, S. 27-34.
- Stürzer, M. (2003a). Geschlechtsspezifische Schulleistungen. In: M. Stürzer, R. Roisch, A. Hunze & W. Cornelißen. *Geschlechterverhältnisse in der Schule*. S. 83-116. Opladen: Leske & Budrich.

Peter Birkel: Geschlechtsspezifische Effekte bei der Zensurenggebung im Fach Mathematik? 16
Version 01.12.2011

- Stürzer, M. (2003b). Unterrichtsformen und die Interaktion der Geschlechter in der Schule. In: M. Stürzer, R. Roisch, A. Hunze & W. Corneließen. *Geschlechterverhältnisse in der Schule*. S. 151ff. Opladen: Leske & Budrich.
- Tent, L. & Birkel, P. (2010). Zensuren. In: D.H. Rost (Hrsg.). *Handwörterbuch Pädagogische Psychologie*. S. 949-958. Weinheim: Beltz.
- 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, S. 101-111.
- vbw – Vereinigung der Bayerischen Wirtschaft e.V. (Hrsg.). (2009). *Geschlechterdifferenzen im Bildungssystem. Jahresgutachten 2009*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Autor: **Dr. Peter Birkel**, PH Weingarten,
Doggenriedstr. 18,
88250 Weingarten
E-Mail: peter.birkel@web.de

Der **Datensatz kann** beim Autor für weitere Verwendungen **angefordert werden**.