

## Peter Birkel

# Gender-spezifische Effekte bei Deutschnoten im Zeugnis?

### *Gender-specific effects when marking the subject German?*

#### **Zusammenfassung:**

3217 Zeugnisnoten im Fach Deutsch aus 5. und 6. Klassen der Haupt- und Realschule wurden auf gender-Effekte geprüft. Es zeigte sich ein deutlicher Effekt auf Seiten der Schülerinnen und Schüler, der auch ein gewisses Maß an Relevanz beanspruchen kann. Ähnlich wie bei den Kompetenzuntersuchungen schnitten Mädchen auch bei den Zensuren besser ab als Jungen. Der gender-Effekt der Lehrkräfte wurde zwar ebenfalls signifikant und wies die Lehrerinnen als mildere Beurteilerinnen aus, aber der Faktor klärte so wenig Varianz, dass man ihm kaum wirkliche Bedeutung beimessen kann. Interessant war auch die signifikante Interaktion zwischen dem Geschlecht der Schülerinnen und Schüler und dem der Lehrkräfte. Sie zeigte, dass die Jungen von Lehrerinnen und Lehrern gleich beurteilt wurden, die Mädchen aber von den Lehrerinnen noch einmal einen Bonus bekamen im Vergleich zu den Beurteilungen durch die Lehrer. Wegen der geringen Varianzklärung darf man aber auch diesem Interaktionseffekt kaum Relevanz zu sprechen, auch wenn Carter (1952) eine ähnliche Interaktion bei Beurteilungen im Fach Algebra fand. Als vergleichsweise starker Faktor erwies sich die Schulartzugehörigkeit. Bei einer Varianzklärung von gut 6% erhielten Realschülerinnen und –schüler bessere Noten im Fach Deutsch als die Hauptschülerinnen und –schüler.

**Schlüsselbegriffe:** Deutschnoten, Schülergeschlecht, Lehrergeschlecht, Schularteffekt

#### **Summary:**

The grades in the subject of German from school reports given to 3217 5<sup>th</sup>- and 6<sup>th</sup>-graders in secondary schools (Haupt- and Realschule) have been checked in order to look for gender effects on both sides, the pupils' and the teachers' side. A significant and partly relevant effect could be shown on the side of the pupils. Similar to the results of studies regarding competence, where girls achieved higher results, it was found that girls also got better marks than boys. The gender effect on the side of the teachers was also significant. Female teachers gave better marks than male teachers, but the relevance of this result was rather low, because that factor showed only a small amount of explanation of variance. A significant interaction was found between the factors "gender of pupils" and "gender of teachers", which was quite interesting. The average grades boys got were exactly the same from both male and female teachers, but the girls got a greater bonus in the grades given by female teachers as compared to male teachers. Although this interaction effect appeared to be in accordance with the findings of Carter (1952), who investigated marks in Algebra, we have to take into account that the relevance of this effect was also slightly low. A strong influencing factor was found to be the type of school pupils attend. With an explained variance of more than 6% pupils from Realschulen obtained better marks than those from Hauptschulen.

**Keywords:** school grades in the subject of German, gender of pupils, gender of teachers, effect of type of school

## 1. Untersuchungsanlass

Die internationalen und nationalen Studien zur Leistungsfähigkeit deutscher Schüler in verschiedenen unterrichtlichen Domänen (PISA, TIMMS, IGLU, LAU, DESI, VERA) erbrachten immer wieder ganz spezielle Leistungsprofile für Jungen und Mädchen. Mädchen zeigen in den sprachlichen Leistungsbereichen, denen auch das Fach Deutsch zuzuordnen ist, eine deutliche Überlegenheit gegenüber den Jungen (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, Klieme et al. 2010), während die Leistungen im mathematisch-naturwissenschaftlichen Bereich im Verlauf der schulischen Entwicklung immer mehr zu einer Domäne der Jungen zu werden scheinen (s.a. Birkel 2011, eingereicht). Nun sind die großen Vergleichsuntersuchungen mit dem Einsatz von Tests gelaufen, die möglichst objektiv Auskunft geben sollten über die Kompetenzen der Schüler in den jeweiligen Fächern. Da aber Zeugnisnoten vielerlei zusätzliche Funktionen erfüllen sollen (vgl. Ingenkamp 1995, Kap. II), ist davon auszugehen, dass es zwar signifikante Korrelationen zwischen der ermittelten Leistungsfähigkeit in Tests und den Zeugnisnoten gibt (vbw 2009), dass aber andererseits auch zum Teil große Diskrepanzen beobachtbar sein könnten. Somit stellte sich die Frage, ob die Überlegenheit der Mädchen in den sprachlichen Bereichen sich auch entsprechend in den von Lehrkräften erteilten Deutschnoten niederschlagen würde.

In der schulischen Entwicklung scheint sich die Überlegenheit der Mädchen in den sprachlichen Unterrichtsbereichen bereits bis zum Ende der Grundschulzeit herauszukristallisieren (Tiedemann & Faber 1994). Die findet auch ihren Niederschlag in den für die Geschlechter erteilten Deutschnoten. So berichtet Giest (1996) über einen Unterschied bei den Zensuren am Ende der 4. Klasse, der 0,8 Notenstufen beträgt (Mädchen 1,9 – Jungen 2,7). In Klasse 6 erhöht er sich auf 0,9 Notenstufen, Mädchen erhalten im Schnitt eine 2,3 und Jungen 3,2. Bis zur 8. Klasse sinkt die Durchschnittsnote für die Mädchen auf 2,5, während die Jungen sich nicht weiter verschlechtern. Die Geschlechtsunterschiede waren sowohl insgesamt, als auch auf den jeweiligen Jahrgangsstufen statistisch hoch signifikant ( $p < .001$ ).

Bereits 1952 führt Carter eine Reihe von empirischen Belegen für die Tatsache an, dass Mädchen auch in der Sekundarstufe I in allen Fächern generell Vorteile gegenüber von Jungen haben (s. S. 148-150). Neuere Untersuchungen sprechen allerdings eher dafür, dass die mathematisch-naturwissenschaftlichen Fächer in der Sekundarstufe eher eine Jungendomäne zu sein scheinen (mit Ausnahme des Faches Biologie), während die sprachlichen und die dem musisch-ästhetischen Bereich zuzuordnenden Fächer eindeutig Mädchendomänen sind (s. z.B. Stürzer 2003a, Dt. PISA-Konsortium 2007, vbw 2009, Birkel 2011 eingereicht). Das wird auch durch die jüngste PISA-Studie erneut bestätigt (Klieme et al. 2010). Die Überlegenheit der Mädchen in den Teilkompetenzen, die zum Bereich des Faches Deutsch gehören, scheint viel größer zu sein, als die der Jungen im mathematisch-naturwissenschaftlichen Bereich, wodurch man die Mädchen in der Sekundarstufe als die insgesamt erfolgreicherer Schüler ansehen kann. Bildungserfolg oder –misserfolg scheint eben doch auch eine Geschlechterfrage zu sein (Crotti 2006).

Erklärungsversuche dazu 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, 2010 eingereicht). Garbe (2011, S. 26) verweist darauf, dass die Erbanlagen oder Hirnstrukturen vermutlich nicht für die Kompetenzunterschiede bei den Geschlechtern verantwortlich gemacht werden können. Die Ursachen seien viel eher im soziokulturellen Bereich zu suchen. Eine zentrale Rolle scheint dabei das Leseverhalten und dessen unterschiedliche Entwicklung bei Jungen und Mädchen zu spielen, das die sprachlichen Kompetenzen deutlich fördern kann. Wallner-Paschon (2010) verweist zudem auf den deutlichen Zusammenhang zwischen der Lesefertigkeit der Schülerinnen und Schüler und den Deutschnoten. Die Rangkorrelation läge bei  $r = .53$  ( $p < .001$ ). Garbe zeigt in ihren Publikationen (2007, 2008, Philipp & Garbe 2007) immer wieder auf, dass die Lesekompetenzen von Jungen und Mädchen sich in der Grundschule noch nicht so stark unterscheiden (Differenz 7 Punkte in der 4. Klasse) und erst in der Sekundarstufe massiv hervortreten (40 Punkte in der 9. Klasse). Sie erklärt das mit einem Rückgang der Lesemotivation bei den Jungen, der darauf zurückzuführen sei, dass Jungen das Lesen als eine typisch „weibliche“ Art erleben, sich die kulturelle und mediale Welt zu erschließen. *„Jungen gewinnen so den Eindruck, dass Lesen ihrer männlichen Identität nicht so zugute kommt. Spätestens in der Pubertät geraten sie in eine ‚Lesekrise‘.“* (Garbe 2011, s.26). Für Jungen scheinen audiovisuelle und digitale Medien viel attraktiver zu sein, zumal sie ihnen mehr Unterhaltung zu bieten und weniger Mühe zu erfordern scheinen. So saßen sie häufiger vor Ihrem Gameboy, dem Fernseher oder der Computer. Diese Tätigkeiten scheinen aber Phantasie und eigene Vorstellungswelt nicht so fördern wie die Entwicklung stabiler Lesegewohnheiten. Hinzu kommt noch die lesebiografische Erkenntnis einer „feminierten Lesekultur“ in der Schule, die durch eine *gendersensible* Leseförderung aufgebrochen werden müsste, in der auch die Bedürfnisse der Jungen eine stärkere Berücksichtigung erfahren könnten. Man darf wohl davon ausgehen, dass die höhere Lesemotivation der Mädchen dazu beiträgt, deren Leistungen im Fach Deutsch zu befördern, was seinen Ausdruck in den zunehmenden Kompetenzvorsprüngen im Bereich der sprachlichen Fähigkeiten in der Sekundarstufe

findet. Zeugnisnoten müssten diese Kompetenzunterschiede abbilden, auch wenn noch viele andere Einflussfaktoren in ihre Festlegung eingehen können. Wie würden sich die Verhältnisse aber in der Orientierungsstufe auf der Ebene der Zeugnisnoten darstellen, also im Übergang von der Grundschule in die Sekundarstufe I? Erklärt das Geschlecht der Schülerinnen und Schüler einen relevanten Anteil an der Notenvarianz im Fach Deutsch?

Über den Einfluss des Geschlechts der Lehrkräfte auf die von ihnen erteilten Zeugnisnoten wissen wir relativ wenig. Beurteilen Lehrer und Lehrerinnen ihre Schüler praktisch gleich oder hängt die Zeugnisnote im Fach Deutsch auch vom Geschlecht der Lehrkraft ab, die dieses Fach unterrichtete? Im Rahmen einer Lehrveranstaltung zu Problemen der Leistungsbeurteilung von Schülern wurde der von Ingenkamp (1995) abgedruckte Beitrag von Robert Scriven Carter aus dem Jahr 1952 kritisch diskutiert. Carter fand heraus, dass die Lehrerinnen im Fach Algebra als Teilbereich der Mathematik deutlich mildere Zensuren gaben als ihre männlichen Kollegen (*Diff* = 6,44 Punkte auf einer 100-Punkte-Skala,  $t_{het} = 5,73$ ,  $df = 196$ ,  $p < .001$ , nachberechnet nach den Originalangaben). Zwar gaben sowohl die Lehrerinnen als auch die Lehrer den Mädchen bessere Zensuren als den Jungen, aber es zeigte darüber hinaus eine Interaktion zwischen dem Geschlecht der Lehrkräfte und dem der Schülerinnen und Schüler. Die Lehrerinnen gaben den Mädchen trotz etwas schwächeren Abschneidens bei einem Test einen Zensurenvorsprung von 4,08 Punkten ( $t_{hom} = 2,27$ ,  $df = 100$ ,  $p < .05$ ) und werteten damit die Jungen noch etwas stärker ab als ihre männlichen Kollegen. Bei ihnen lag der Vorsprung der Mädchen bei nur 2,89 Punkten auf der 100-Punkte-Skala ( $t_{het} = 2,25$ ,  $df = 131$ ,  $p < .05$ ). Trotz der stärker ausgeprägten Bevorzugung der Geschlechtsgenossinnen erteilten die Lehrerinnen den Jungen immer noch im Schnitt bessere Noten als die Lehrer.

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 eine Untersuchung, die fast 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 (2010 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 aber 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 im Fach Deutsch

überhaupt auswirkt, sollte nach der Auffassung der Studierenden in einer Untersuchung unter heutigen Bedingungen einmal erneut geprüft werden.

Im Hinblick auf die Beurteilungsgerechtigkeit der Lehrkräfte bezweifelten vor allem die weiblichen Studierenden, dass auch heute noch die Lehrerinnen generell mildere Zensuren erteilen würden im Vergleich zu ihren männlichen Kollegen. Ein wichtiger Grund für diese Zweifel war vor allem die Vermutung, dass es doch im Fach Deutsch auch möglich sein sollte, Schülerleistungen einigermaßen objektiv, d.h. unabhängig von der bewertenden Person, zu beurteilen. Zwar wisse man um die Problematik der Aufsatzbeurteilung (s. Birkel 2003), aber es bildeten doch auch andere Leistungen, wie z.B. das vermeintlich objektiver zu beurteilende Diktat (Die Untersuchung von Birkel (2009) war den Studierenden noch nicht bekannt.) die Grundlage für die Notengebung im Zeugnis. Wenn eine objektiv feststellbare Anzahl von Fehlern vorliege, dürfte das Geschlecht der beurteilenden Lehrkräfte kaum eine Rolle spielen. 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 Interaktion 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 diese generelle Tendenz zu milderer Beurteilung ihren Ursprung haben könnte

- a) im *Wesen der Weiblichkeit* 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 weibliche Sozialisation und damit als immer wieder beobachtbar zu betrachten. Würde dieser Unterschied nicht wieder auftreten, käme als Erklärung die inzwischen eher 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 genereller 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 Zensurenggebung 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 (1989) beschreibt damit sicher eine vielfach vermutete Erklärung der Urteilsunterschiede, liefert aber keinen empirischen Beweis für die Richtigkeit ihrer Interpretation, da Ihre

Erkenntnisse auf Interviews mit wenigen Lehrerinnen beruhen. Die Ausführungen von Flaake (1989) legen die Vermutung nahe, dass eine positivere Beurteilung von Schülerleistungen eine typisch weibliche Eigenschaft sei. Es wäre dann also zu erwarten, dass diese Tendenz einen relevanten Einfluss auf die Zensurengebung durch die Lehrerinnen haben müsste. Von daher erscheint es einmal mehr als sinnvoll und notwendig, das Urteilsverhalten der Lehrkräfte unter heutigen Bedingungen im Fach Deutsch bei einer großen Stichprobe empirisch neu zu untersuchen, zumal die Erwartungen der Studierenden eher im Gegensatz zu Flaakes Überlegungen stehen.

In den letzten Jahren wurde vor allem von Politikern und Wirtschaftsvertretern gern die Feminisierung des Lehrberufes mitverantwortlich gemacht für die geringeren Leistungen der Jungen in der Schule generell und im Fach Deutsch speziell ([www.frauensicht.ch](http://www.frauensicht.ch) 2004, [www.linzerin.at](http://www.linzerin.at) 2004). Besonders die konservativen Politiker in einigen europäischen Ländern scheinen geneigt zu sein, den „vorherrschend weiblichen Werten“ in den Schulen die Schuld für das schwächere Abschneiden der Jungen zu geben. Dänische Gymnasialrektoren befürchten gar gesellschaftliche Probleme, wenn nicht eine gezielte Förderung der Jungen einsetze, da dort bereits bis 75% der Gymnasiasten weiblichen Geschlechts seien. Ohne genauere Quellenangabe wird auf diesen Internetseiten auf Untersuchungen der Universität Nijmegen verwiesen, denen zufolge das Geschlecht der Lehrkraft für die Leistung der Schüler unwichtig zu sein scheint. Untersucht wurde die Leistung von etwa 5000 Schülerinnen und Schülern. Zwar seien die Schulleistungen der Mädchen im Schnitt besser als die der Jungen, aber in der Sekundarstufe seien in den Niederlanden die Lehrerinnen nicht über-, sondern untervertreten. Es gebe bis heute keinen seriösen wissenschaftlichen Beweis für einen Zusammenhang zwischen den Schulleistungen und den Geschlecht der Lehrkräfte. Auf der Internetseite der Schweizer Zeitschrift „Frauensache“ wird Georg Stöckli vom PI der Uni Zürich mit der Erklärung zitiert, die schlechteren Leistungen (und damit auch die Zensuren?) der Jungen seien eher auf die Tendenz zur Überschätzung der eigenen Fähigkeiten und der daraus resultierenden geringeren Lernbereitschaft zurückzuführen. Hinzu kämen nach der Ansicht des Schweizer Psychologen Guggenbühl Faktoren wie die an den Mädchen orientierten Arbeitsmethoden, die Idealisierung der Harmonie oder die Vermeidung von Wettbewerbssituationen, die die Jungen eher ansprechen könnten. Wenn man diesen Publikationen folgt, wäre zu erwarten, dass es im Gegensatz zu den bei Carter (1952) zitierten Autoren und zu Klein (2004) bei der Analyse von Zeugnisnoten in der Orientierungsstufe keinen nennenswerten Zusammenhang zwischen dem Lehrergeschlecht und den erteilten Zeugnisnoten gibt. Damit wäre ein weiterer Grund gegeben, hier einmal erneut auf mögliche Zusammenhänge zu achten.

Besonders im Zusammenhang mit der Frage, warum die Leistungen der Jungen in der Schule so stark hinter denen der Mädchen zurück bleiben, wurde immer wieder der Verdacht geäußert, dass die Feminisierung der Schule daran Schuld tragen könnte oder zumindest ihren Anteil daran hätte (Arnot et al. 1998, Budde 2008, Dee 2007, Diefenbach & Klein 2002, Driessen 2007, Horstkemper 1999, Preuss-Lausitz 2005, Rose & Schmauch 2005). Damit sind zwei Aspekte angesprochen: 1. Die Tatsache, dass der Anteil weiblicher Lehrkräfte ständig steigt, und 2. dass sich Unterrichtsinhalte und Schulkultur feminisiert haben und von daher eher den Erwartungen und Bedürfnissen der Mädchen entspricht. Tatsächlich sei der Anteil weiblicher Lehrkräfte an den allgemeinbildenden Schule von 1965 bis 2007 von 45,8% auf 68,5% gestiegen (Helbig 2010, S.96). Dass sich aber allein der Anteil weiblicher Lehrkräfte negativ auf das Leistungsverhalten der Jungen auswirkt, ist bisher weder national noch international nachgewiesen (Budde 2008, Driessen 2007, Stamm 2008, vbw 2009). Bei der Feminisierung der Schulkultur führt Helbig (2010, S. 97f) folgende Argumente auf, die aber noch immer Hypothesen sind:

1. In der Schule gelten weibliche Normen, die durch Frauen bestimmt wurden. Denen müssen sich Jungen anpassen (Sexton 1969, S. 10f). „Richtigen Jungs“ oder „männlichen Typen“

fällt es wesentlich schwerer diese zu internalisieren als den Mädchen (Martino 2008, S. 206).

2. Das Fehlen männlicher Vorbilder und Identifikationsfiguren wirkt sich negativ auf Motivation und Leistung der Jungen aus.
3. Die Überbewertung typisch weiblicher Mädchenqualitäten durch Lehrerinnen führt zu einer Unterbewertung der Leistungen der Jungen. Sie bekommen bei gleichen Leistungen schlechtere Beurteilungen (Stamm 2008).
4. Die Schulkultur ist unabhängig vom Geschlecht der Lehrkraft „weiblich“ ausgerichtet und spricht stärker Lernstil und Lernbereitschaft der Mädchen an (Bacher et al. 2008).

Helbig (2010, S.107) geht in seiner Untersuchung diesen Fragen nach und kommt zu dem Ergebnis, dass der Anteil an Lehrerinnen weder die Kompetenzentwicklung der Jungen beim Lesen noch in der Mathematik maßgeblich beeinflusst, sie aber zumindest im Mathematik etwas schlechtere Noten bekommen je mehr Lehrerinnen an einer Schule unterrichten. Daraus aber eine bedenkliche Benachteiligung der Jungen durch die Lehrerinnen abzulesen sei vermutlich überzogen. Bei gleicher Kompetenz bekommen Jungen sowohl in Deutsch als auch in Mathematik schlechtere Noten. Das hat auch Auswirkungen auf die Übergangsempfehlungen und führt dazu, dass Jungen seltener eine Gymnasialempfehlung erhalten an Schulen mit hohem Frauenanteil. Fraglich bleibt allerdings die Aussagekraft des Prozentanteils an weiblichen Lehrkräften auf Schulebene bezüglich der Kompetenzentwicklung und deren Beurteilung, weil die individuelle Passung zwischen dem Geschlecht der Lehrkraft und dem der Schülerinnen und Schüler nicht berücksichtigt wird. Wenn ein Schüler in einem Fach gar nicht von einer Lehrerin unterrichtet wurde, könnte sich der Frauenanteil am Kollegium höchstens indirekt auswirken über das Schulklima und die –kultur. Damit wird für unsere Untersuchung deutlich, dass man keine eindeutigen Zusammenhänge zwischen dem Geschlecht der Lehrkräfte und der Beurteilung der Schulleistungen der Schülerinnen und Schüler postulieren kann. Wohl aber könnte es eine Interaktion zwischen dem Geschlecht der Lehrkräfte und dem der Schülerinnen und Schüler bezüglich der Leistungsbeurteilung geben.

Die Frage nach der Abhängigkeit der Schülerbeurteilungen von der Interaktion zwischen dem Geschlecht der Lehrkräfte und dem der Schülerinnen und Schüler wurde in neuerer Zeit angeheizt durch Untersuchungsergebnisse, die von Diefenbach & Klein (2002) publiziert wurden. Sie fanden heraus, dass es offenbar eine solche Interaktion gibt, wie sie in ähnlicher Form schon von Carter (1952) beschrieben wurde. Lehrerinnen gewährten ihren Geschlechtsgenossen bessere Noten als den gegengeschlechtlichen Schülern. Als Erklärung wurde aufgeführt, dass Lehrerinnen „gegenüber den Eigenarten von Jungen weniger tolerant und bei einigen sogar behindernd wirkten, wie z.B. hinsichtlich des größeren Bewegungsdranges der Jungen. Die Haltung der Lehrerinnen bewirke

1. schlechtere Noten von Jungen,
2. eine geringere Häufigkeit bei Empfehlungen für „höhere“, nämlich Real- oder gymnasiale weiterführende Schulen.“ (Balluseck, 2009, S.1)

Ergebnisse der IGLU-Studie scheinen die Ergebnisse von Diefenbach & Klein (2002) zu stützen. Dort wird festgestellt, dass die schlechteren Noten der Jungen als „systematische Ungleichbehandlung der Jungen durch die überwiegende Zahl der Grundschullehrerinnen interpretiert werden“ könne. (Valtin et al. (2005, S. 226)

Diese Befundlage erscheint als Rechtfertigung für die Diskussion um die „Jungen als Verlierer“ (Balluseck, 2009), wie sie in Medien und Politik seit einiger Zeit geführt wird. Dabei dreht sich die Diskussion nach Balluseck (2009) im Wesentlichen um zwei Behauptungen:

- „Weibliche Lehrkräfte gäben Jungen bei gleichen Leistungen schlechtere Noten. Diefenbach ist in ihrer neuesten Veröffentlichung von dieser Behauptung abgerückt, denn es gibt keine wissenschaftlichen Beweise für diese Behauptung. Faktisch „unterliegen weib-

*liche Klassenlehrer bei ihrer Notenvergabe in geringerem Maße Verzerrungen" als männliche (Schultheis 2008; vgl. auch Kuhn 2008). In einem Telefonat am 15.12.08 versicherte mir auch der Leiter der Studien IGLU und TIMMS, Wilfried Bos, dass es keinerlei Belege dafür gebe, dass weibliche Lehrkräfte Jungen schlechter beurteilen als männliche. Tatsache sei allerdings, dass Jungen es generell - also bei weiblichen und männlichen Lehrkräften - schwerer hätten, bei gleichen Leistungen die gleichen Noten wie Mädchen zu erhalten. Damit ist der Behauptung von Valtin u. a. 2005 in der IGLU-Studie der Boden entzogen.*

- *Jungen störten häufiger den Unterricht, ihr externalisierendes Verhalten sei Anlass für die Lehrerinnen, sie schlechter zu bewerten. Diefenbach (2008) hat daraufhingewiesen, dass das häufig externalisierende Verhalten von Jungen in den Bildungsinstitutionen weniger „passt" als das eher angepasste Verhalten der Mädchen. Daher werden sie häufiger als Mädchen als störend wahrgenommen. Die Annahme, dass weibliche Lehrkräfte auf dieses Verhalten durch schlechtere Noten überreagieren, lässt sich jedoch nicht erhärten, im Gegenteil: Weibliche Lehrkräfte versuchen im Vergleich zu ihren männlichen Kollegen stärker, „ihre Notenvergabe nicht von disziplinarischen Wahrnehmungen des schülerischen Unterrichtsverhaltens beeinflussen zu lassen." Sie konzentrieren sich bei ihrer Notenvergabe in erster Linie an der Erfüllung von Hausaufgaben und Fleiß (Schultheis 2008, S. 369). Schülerinnen und Schüler, die von weiblichen Lehrkräften unterrichtet werden, haben außerdem eine „deutlich positivere Einstellungen zur Schule" (Rohrmann 2008 b, S. 115). Dies könnte schwerlich der Fall sein, wenn sie sich ungerecht behandelt fühlten.“ (Balluseck, 2009, S1/2)*

Die Entwicklung der Geschlechtsidentität geschieht in wesentlichen Bestandteilen durch die Sozialisation im Elternhaus. Dabei kommt es auf die Freiräume an, die den Kindern bei der Findung des eigenen Geschlechts eingeräumt werden. Rigide Geschlechtnormen in der Erziehung begünstigen bei Jungen abwertende Haltungen gegenüber allem „Weiblichen“ (Balluseck 2009). Diese Jungen bekämen nach Meinung von Balluseck (2009) Probleme, sobald sie in der Schule mit Lehrerinnen konfrontiert werden, weil die in der Schule erlebte weibliche Professionalität und Kompetenz in krassem Gegensatz zu der Rolle der Frau innerhalb der Familie steht. Solche Gegensätze seien besonders oft beobachtbar in Migrantenfamilien und problembeladenen Familien (z. B. Erwerbslosigkeit, soziale Schwäche, geringe Bildung). Das professionelle Handeln wird durch solche Einflüsse vermutlich besonders erschwert für Lehrerinnen. Ob aber deshalb Lehrerinnen die Zensuren z.B. als Disziplinierungsinstrumente gegenüber den Jungen einsetzen und sie darum schlechter beurteilen, erscheint doch als fraglich bis unwahrscheinlich (s. Zitat oberhalb) und wurde von den überwiegend weiblichen Studierenden unserer Hochschule heftig in Zweifel gezogen.

Damit scheint aber die Frage nach wie vor offen, ob denn nun wirklich die Lehrerinnen diejenigen sind, die für die schlechteren Noten der Jungen verantwortlich sind. Auch der andere Aspekt muss in Betracht gezogen werden: Vielleicht bilden die Lehrerinnen nur den tatsächlichen Leistungsvorsprung der Mädchen adäquat ab und die Lehrer ignorieren ihn, wenn sie Mädchen und Jungen relativ ähnliche Beurteilungen im Fach Deutsch zukommen lassen. Aber selbst signifikante Ergebnisse, die die positivere Beurteilung der Mädchen durch die Lehrerinnen zu stützen scheinen, müssten allerdings auf ihre Relevanz geprüft und mit anderen Faktoren verglichen werden, die ebenfalls einen Einfluss auf die Beurteilung der Schülerinnen und Schüler haben können. Wenn der Faktor „Geschlecht der Lehrkraft“ nur wenig Varianz klärt im Vergleich zu anderen Faktoren, wäre er möglicherweise doch eher zu vernachlässigen. Das kann vor allem bei Untersuchungen mit großen Stichproben leicht vorkommen.

Eine weitere Kontroverse ergab sich bei der Frage, inwieweit sich die Zensuren von Haupt- und Realschülern in der Orientierungsstufe gleichen könnten. Einerseits wurde im Se-

minar diskutiert, dass die bessere Leistungsfähigkeit der Realschüler ihren Niederschlag in den Zensuren finden könnte, andererseits wurde dagegen gehalten, dass die Lehrkräfte keinen direkten Vergleich zwischen Haupt- und Realschülern anstellen konnten und von daher sich bei der Notenfindung eher an der durchschnittlichen Leistungsfähigkeit ihrer jeweiligen Klasse orientieren würden. Die von Ingenkamp (1995, S. 194ff, s. auch Ingenkamp & Lissmann 2005) durchgeführte Untersuchung zur Vergleichbarkeit von Zensuren verschiedener Klassen legte nahe, dass unabhängig vom Leistungsniveau der Klassen die Lehrkräfte bei der Vergabe der Zensuren in aller Regel die Notenskala ausschöpften ohne Berücksichtigung der möglicherweise unterschiedlichen durchschnittlichen Leistungsfähigkeit der Schülerinnen und Schüler. Nur so konnte es geschehen, dass Schülerinnen und Schüler mit gleicher, durch einen Test gemessenen Leistung in einer Klasse die Note „sehr gut“ erhielten, während sie im Kontext einer anderen Klasse ebenso gut die Note „mangelhaft“ hätten erhalten können. Auch Schrader & Helmke (2001) bescheinigen den Lehrkräften, dass sie bei ihrer Notengebung vornehmlich einen sozialen Bezugsmaßstab zugrunde legen und Schülerinnen und Schüler im Vergleich zu Leistungen der anderen innerhalb der Klasse beurteilen. Unterstellt man diese Beurteilungstendenz, so wäre zu erwarten, dass die Noten der Haupt- und Realschülerinnen und –schüler sich ebenso wenig unterscheiden wie die Noten der Schüler in den untersuchten Jahrgangsstufen.

Auf der anderen Seite erscheinen immer wieder die Kompetenzunterschiede in den zum Bereich des Faches Deutsch gehörenden Teilkompetenzen wie z.B. Lesen oder Rechtschreiben zwischen den Schularten deutlich zu Tage zu treten (PISA, LAU, DESI, VERA). Realschülerinnen und –schüler schneiden dabei deutlich besser ab als die Hauptschülerinnen und –schüler. Diese Unterschiede zeigen sich allerdings in den schulartübergreifend einsetzbaren Tests. Ob jetzt allerdings Lehrkräfte an der jeweiligen Schulart diese Unterschiede ermessen können und entsprechend bei der Zensurenvergabe berücksichtigen, ist nicht bekannt. Wenn dem so wäre, könnten sich die Durchschnittszensuren an den beiden Schularten unterscheiden. Ein weiteres Argument in dieser Richtung könnte der Hinweis auf die ungleiche Geschlechterverteilung in den Schularten sein. Der stärkere Jungenanteil an der Hauptschule könnte generell die Durchschnittsnoten an dieser Schulart belasten. Insofern könnte es als u.U. ertragreich angesehen werden, Schulart und Jahrgangsstufe mit zu untersuchen.

## **2. Untersuchungsansatz**

Die Zeugnisnoten am Ende der Klassen 5 und 6 an Haupt- und Realschulen im Regierungsbezirk Südwürttemberg<sup>1</sup> wurden in der Form erfasst, dass ohne Kenntnis der Namen von Schülerinnen, Schülern und Lehrkräften oder der 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. Die Anonymisierung der Daten bezüglich Schul- und Klassenzugehörigkeit war Voraussetzung für die Genehmigung der Untersuchung.

Mit der Festlegung der Zeugnisnote als abhängiger Variable weichen wir allerdings von der Untersuchung von Carter (1952) insofern ab, als wir ein globales Fähigkeitsmaß zur Analyse heranziehen wollten, in das vielerlei Faktoren, und damit auch Faktoren der Persönlichkeit von Schülerinnen, Schülern und Lehrkräften, eingehen (vgl. Ingenkamp 1995, S. 45ff). 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 Leistungsfeststellung haben. Tests wurden daher zur möglichst objektiven Beurteilung der Leistungsfähigkeit von Schülerinnen und Schülern eingesetzt, wie z.B. in den TIMMS-, PISA-, LAU-, DESI oder IGLU-Untersuchungen, bei denen es auch um die Erfassung von eventuellen Geschlechtsunterschieden auf Seiten der Schülerinnen und Schüler ging. Wie sehr Testergeb-

---

<sup>1</sup> Die Genehmigung zu dieser Untersuchung wurde von Oberschulamt in Tübingen erteilt.



nisse und Zensuren voneinander abweichen können, darauf weist bereits Carter (1952) hin. Die PISA-Ergebnisse belegten allerdings noch klarer, wie sehr sich Testleistungen und Zensuren von Schülerinnen und Schülern 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

- a) *gender*-Einflüsse auf Seiten der Schülerinnen und Schüler im Fach Deutsch sich in den Zensuren ebenso widerspiegeln könnten wie in den Testergebnissen 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.

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 auf Gleichverteilung der Geschlechter bei den Haupt- und Realschülerinnen und -schülern ergab einen  $\chi^2$ -Wert<sub>df=1</sub> 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 47,8% aller Zensuren von Lehrern erteilt worden waren, die das Fach Deutsch unterrichteten, und 52,2% der Zensuren von Lehrerinnen. An den Realschulen lag die Geschlechterverteilung bei 46,2% ♂ zu 53,7% ♀. Eine Überprüfung der Gleichverteilung des Beurteilergeschlechts bei Haupt- und Realschulen mit dem  $\chi^2$ -Test ergab einen insignifikanten  $\chi^2$ -Wert<sub>df=1</sub> von 0,693. Man kann also feststellen, dass an der Haupt- und Realschule in etwa gleich viele Zensuren von Lehrern und Lehrerinnen erteilt worden waren. Dabei war sichtbar, dass insgesamt im Fach Deutsch offenbar mehr Lehrerinnen die Noten erteilt hatten als Lehrer. Eine Überprüfung auf Gleichverteilung der Lehrkraftgeschlechter insgesamt ergab einen  $\chi^2$ -Wert<sub>df=1</sub> von 10,63 ( $p < .01$ ). Dieser Wert signalisiert, dass unabhängig von der Schulart die Feststellung trägt, dass deutlich mehr Lehrerinnen das Fach unterrichtet haben als Lehrer. Das Fach Deutsch scheint auch auf Seiten der Lehrkräfte zu einer Domäne der Frauen zu werden. Es spricht offenbar bereits bei der Studienwahl aufgrund ihrer größeren Literalität mehr Frauen als Männer an. Zwar studieren an der Pädagogischen Hochschule generell schon wesentlich mehr weibliche Studierende (geschätzte 85%), aber im Studienfach Deutsch liegt dieser Prozentsatz im Bereich von 90%.

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 Ordinalskalengualität<sup>2</sup> hat. Bei der Anwendung der Varianzanalyse zur zufallskritischen Beurteilung von

---

<sup>2</sup> 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 Intervallskalenqualität handelt.

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 der der Schülerinnen und Schüler und D = Geschlecht Lehrkraft erweitern. Abhängige Variable war die Zeugnisnote, die verordnungsgemäß 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 152 in der Klasse 6, während bei den Jungen mit 295 Personen die höchste Zellenbesetzung erreicht wird. Trotz dieser Unterschiede sind aber alle Zellen hinreichend häufig besetzt.

A Schulart	Hauptschule								Realschule							
B Klasse	Klasse 5				Klasse 6				Klasse 5				Klasse 6			
C Schülergeschlecht	Junge		Mädchen		Junge		Mädchen		Junge		Mädchen		Junge		Mädchen	
D Lehrer*in geschlecht	L♂	L♀	L♂	L♀	L♂	L♀	L♂	L♀	L♂	L♀	L♂	L♀	L♂	L♀	L♂	L♀
N=	198	246	171	198	213	295	152	176	161	167	178	253	173	199	188	203

Tabelle 1: Varianzanalytisches Design und Zellenbesetzungen

### 3. Hypothesen

- a) Im Einklang mit den Ergebnissen der PISA- und IGLU-Studien wird eine *Überlegenheit der Mädchen gegenüber den Jungen im Fach Deutsch* erwartet. Diese Überlegenheit sollte sich auch in erteilten Zensuren widerspiegeln (Haupteffekt Faktor C).
- b) Dass das Lehrkraftgeschlecht ganz allgemein zu unterschiedlichen Zeugnisnoten im Fach Deutsch führt, wird nicht nur von den Studierenden bezweifelt, sondern auch in den auf den Internetseiten von „Frauensicht“ und „Linzerin“ (2004). Carter (1952) dagegen berichtet im Bereich Algebra von deutlich mildereren Beurteilungen durch die Lehrerinnen. Die neuere Untersuchung von Klein (2004) bestätigt dieses Ergebnis. Die Überlegungen von Marsh et al. (2008) sprechen mit ihrer Invarianzhypothese eher für eine Angleichung der Lehrkrafturteile. Es ist somit *unklar, ob aber mit einer mildereren Zensurenvergabe durch die Lehrerinnen gerechnet werden muss*. Von daher neigen wir hier zur *Nullhypothese* (Haupteffekt Faktor D).
- c) Da in neuerer Zeit (Diefenbach & Klein 2002, Valtin et al. 2005, Helbig 2010) von einer Bestätigung der von Carter (1952) beobachteten Interaktion zwischen dem Geschlecht der Lehrkräfte und dem der Schülerinnen und Schüler bei der Zensurenvergabe berichtet wurde, erschien es uns als Gebot der Vorsicht, hier folgende Hypothese zu formulieren: Es gibt *eine Interaktion zwischen dem Geschlecht von Lehrkräften und Schülerinnen und Schülern* (Interaktion C\*D) in der Form, dass die *Lehrerinnen Mädchen stärker positiv beurteilen* als die Lehrer.
- d) Da Lehrkräfte ihre 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 an Haupt- und Realschulen erteilten Noten* erwartet (Haupteffekt Faktor A).

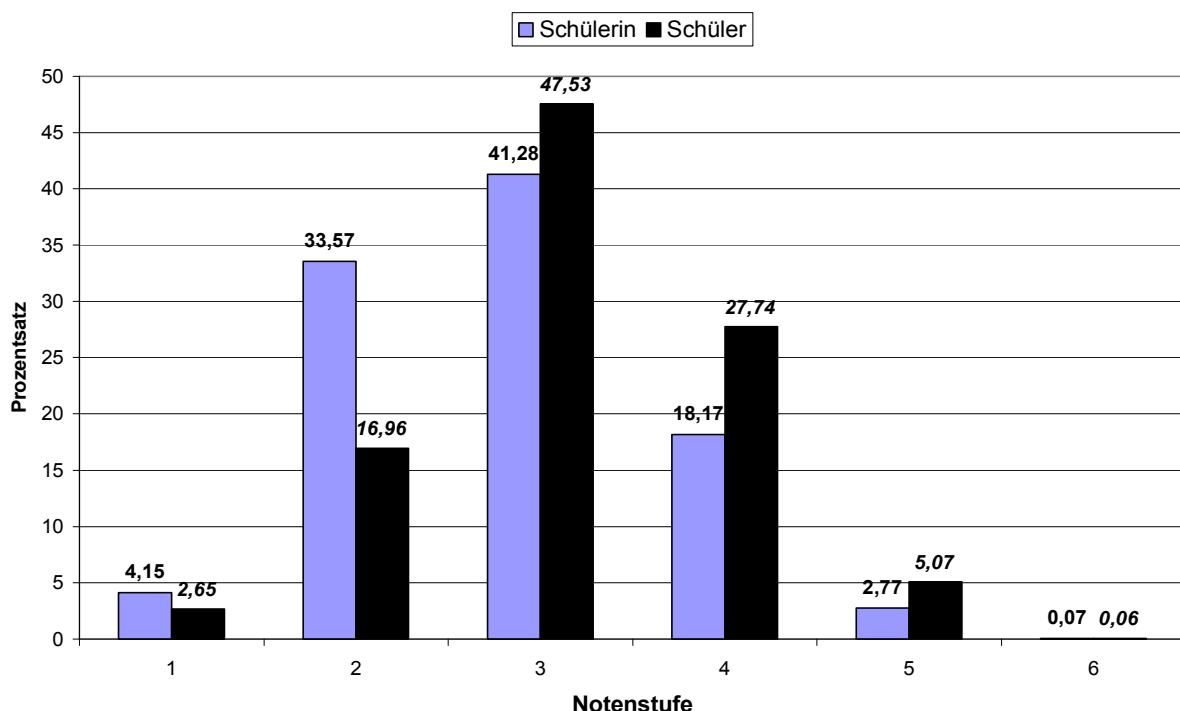
- e) Eine generelle *Veränderung der in den Klassenstufen 5 und 6 erteilten Noten* wird aus dem gleichen Grund *nicht erwartet* (Haupteffekt Faktor B).

## 4. Ergebnisse

### 4.1 Haupteffekte

Mit einem Wert  $F_{1,3201} = 102,63$  wird für den Faktor C = *Schülergeschlecht* eine Signifikanz auf dem 1%-Niveau angezeigt. Jungen erreichen mit einem Mittelwert von 3,11 Beurteilungen, die deutlich schlechter liegen als die der Mädchen mit einem Mittel von 2,81. Die Zensuren der Mädchen liegen also um fast eine Drittel Notenstufe besser als die der Jungen. Unsere Hypothese a) kann damit als bestätigt angesehen werden. Dem Geschlechtsunterschied darf auch durchaus Relevanz zugemessen werden, denn die Varianzklärung von 2,9% kann nicht unbeachtet bleiben.

Abbildung 1: Prozentuale Notenverteilung bei Schülerinnen und Schülern



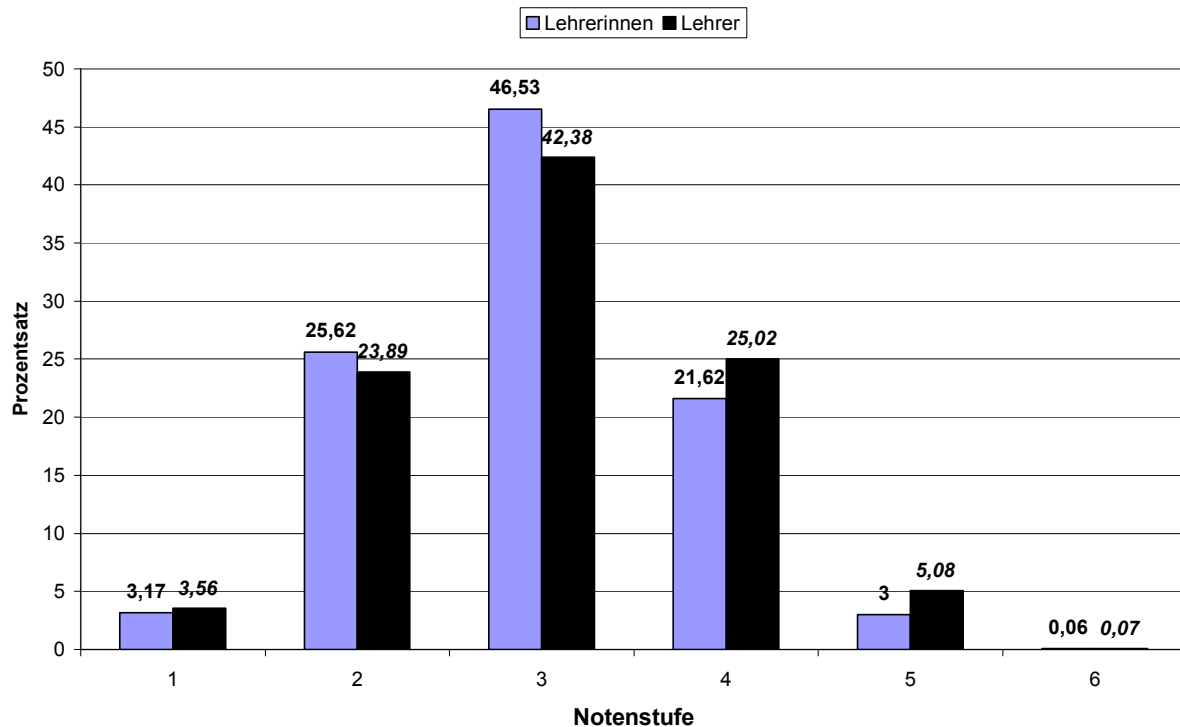
Aus Abbildung 1 ist zu erkennen, dass die Mädchen etwas häufiger die Note „sehr gut (1)“, aber mit 33,57% fast doppelt so häufig die Note „gut (2)“ erhielten im Vergleich zu den Jungen (16,96%), während die Jungen an beiden Schularten häufiger die Noten 3 bis 5 bekamen.

Bei der Beurteilung der *Mittelwertsunterschiede zwischen den von Lehrerinnen und Lehrern erteilten Noten* ergibt sich eine Signifikanz auf dem 5%-Niveau ( $F_{(1,3201)} = 4,95$ ;  $p < .05$ ;  $Eff\%^3 = 0,14$ ). Ähnlich wie bei Carter (1952) und den dort zitierten Untersuchungen und bei Klein (2004) urteilten hier die Lehrerinnen ( $M = 2,93$ ) einen Hauch milder als die Lehrer ( $M = 2,99$ ). Wegen der geringen absoluten Mittelwertsdifferenz und der geringen Varianzklärung wird man aber diesen Unterschied wohl eher vernachlässigen und davon ausgehen können, dass hier kein wirklich bedeutender Beurteilungsunterschied in Abhängigkeit vom Lehrergeschlecht besteht. Die deutlichen Beurteilungsunterschiede, die sich bei Carter (1952) und Klein (2004) finden, sind hier nicht beobachtbar. Man wird daher wohl eher die Nullhypothese b) beibehalten dürfen. Es spricht somit einiges dafür, sich der Invarianzhypothese von Marsh et al. (2008)

<sup>3</sup> Mit dem *Eff%*-Wert gibt Kleiter (2004) die Effektstärke, also den prozentualen Anteil der geklärten Varianz an.

anzuschließen, die eventuell vorhandene Unterschiede vor allem dann für unbedeutend halten, wenn die Varianzklärung – wie jetzt hier – sehr gering ist.

Abbildung 2: Prozentuale Notenverteilung bei Lehrerinnen und Lehrern



Der ohnehin geringe Unterschied bei den von Lehrerinnen und Lehrern erteilten Noten zeigt sich – wie aus Abbildung 1 zu ersehen – noch am ehesten bei den Noten 3 und 4. Lehrerinnen gaben häufiger ein „befriedigend (3)“ ( $Diff = +4,15\%$ ), während die Lehrer bei der Note „ausreichend (4)“ die Nase vorn hatten ( $Diff = -3,4\%$ ). Nicht so deutlich sind die Unterschiede bei der Vergabe der Noten „gut (2)“ ( $Diff = +1,73\%$ ) und „mangelhaft (5)“ ( $Diff = -2,08\%$ ). Bei der Vergabe der Noten „sehr gut (1)“ und „ungenügend (6)“ spielen die Unterschiede praktisch keine Rolle.

Das Ergebnis der Varianzanalyse signalisiert, dass erstaunlicherweise der Faktor A = *Schulart* den stärksten Effekt verursacht ( $F_{1,3202} = 219,93$ ;  $p < .001$ ,  $Eff\% = 6,19$ ). Insgesamt gesehen erhielten die Hauptschülerinnen und -schüler mit einer Durchschnittsnote von 3,18 eine schlechtere Beurteilung als die Realschülerinnen und -schüler, bei denen der Mittelwert bei 2,74 lag. Damit ist unsere Hypothese d) abzulehnen. Möglicherweise spielt zumindest das Wissen um die generelle Überlegenheit der Realschülerinnen und -schüler bei deren Beurteilung in der Orientierungsstufe doch eine gewisse Rolle. Zusätzlich wurde in einer Kovarianzanalyse mit drei Faktoren das Geschlecht der Schülerinnen und Schüler als Kovariate eingegeben. Die Veränderung der F-Wertes ist mit 216,12 ( $p < .001$ ;  $Eff\% = 6,07$ ) ist nur marginal. Damit lässt sich ausschließen, dass das Schülegeschlecht als Moderatorvariable den Haupteffekt des Faktors *Schulart* wesentlich beeinflusst hat.

Schließlich erreicht der Haupteffekt des Faktors B = *Klassenstufe* die Signifikanzgrenze nicht ( $F_{(1/3201)} = 0,51$ , *n.s.*). Man kann also davon ausgehen, dass sich die Beurteilung aller Fünftklässlerinnen und -klässler nicht der durchschnittlichen Beurteilungen der Sechstklässlerinnen und -klässler unterscheidet.

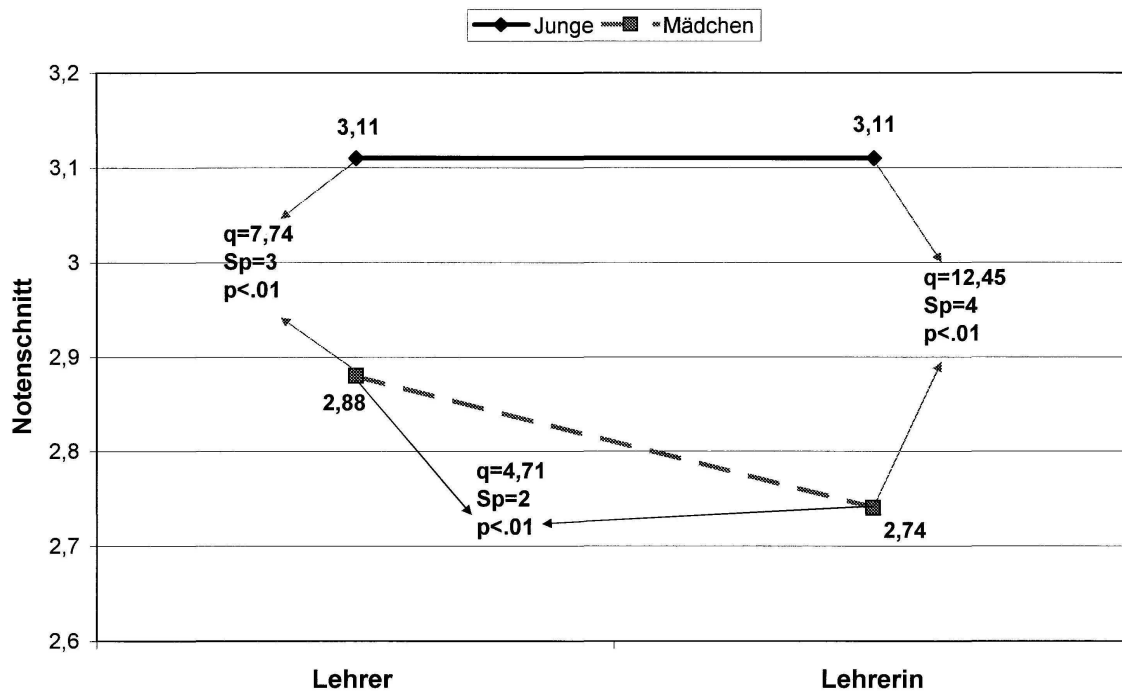
#### 4.2 Interaktionseffekte

Die Befundlage bei Carter 1952, Diefenbach & Klein 2002, Valtin et al. 2005 und Helbig 2010 bezog sich auf die hier untersuchte Interaktion zwischen den Faktoren C (*Geschlecht der*

Schülerinnen und Schüler) und D (*Geschlecht der Lehrkräfte*). Unsere Varianzanalyse weist hier mit einem Wert  $F_{(1/3201)} = 5,54$  eine Signifikanz auf dem 5%-Niveau aus.

Jungen bekommen hier von Lehrern und Lehrerinnen im Schnitt die gleichen Beurteilungen, beide Geschlechter geben den Jungen eine 3,11. Mädchen dagegen werden von Lehrerinnen und Lehrern besser beurteilt als die Jungen, aber die Lehrerinnen geben positivere Urteile ab ( $M = 2,74$ ) als die Lehrer ( $M = 2,88$ ,  $q = 4,71$ ;  $p < .01$ ;  $Sp = 2$ ). Obwohl also andeutungsweise hier ein ähnliches Ergebnis wie bei Carter 1952, Diefenbach & Klein 2002 und Valtin et al. 2005 auftritt, darf man die Signifikanz dieser Interaktion sicher nicht überinterpretieren. Der Effekt ist bei einer Varianzklärung von  $Eff\% = 0,16$  als wenig relevant zu bezeichnen. Man wird also nicht gerade eine unangemessene Übervorteilung der Mädchen durch die Lehrerinnen aus diesen Ergebnissen ablesen dürfen. Bei der genaueren Analyse der Beurteilungsunterschiede zeigt sich, dass sowohl die Lehrer als auch die Lehrerinnen den Mädchen bessere Zensuren erteilen als den Jungen. Allerdings ist der Notenvorteil der Mädchen bei den Lehrerinnen signifikant größer als bei den Lehrern ( $q = 4,71$ ;  $p < .01$ ). Sie erhalten bei den Lehrerinnen im Notenschnitt noch einmal einen Bonus von 0,14 Notenstufen gegenüber den Lehrern.

Abbildung 3: Interaktion zwischen dem Schüler- und Lehrergeschlecht



Auch die Dreierinteraktion zwischen den Faktoren B\*C\*D (*Klassenstufe\*Schülergeschlecht\*Lehrkraftgeschlecht*) erreicht ein Signifikanzniveau ( $F_{(1/3201)} = 4,10$ ;  $p < .05$ ;  $Eff\% = 0,12$ ), ist aber ebenso wenig relevant wie die Zweierinteraktion C\*D. Sie macht aber deutlich, dass der Schereneffekt zwischen der Beurteilung der Jungen und Mädchen durch Lehrerinnen und Lehrer in der 6. Jahrgangsstufe besonders deutlich hervortritt. Geben Lehrer den Mädchen hier eine um 0,17 Notenstufen bessere Beurteilung, so steigert sich die positivere Beurteilung durch die Lehrerinnen auf 0,43 Notenstufen. Im Vergleich zu den Lehrern werden die Jungen schlechter, die Mädchen aber deutlich viel besser beurteilt. Aufgrund der hohen Stichprobenzahl werden diese Unterschiede zwar signifikant, dürfen aber – wie gesagt – wegen der geringen Effektgröße nicht überinterpretiert werden.

Trotz jeweils geringer Relevanz muss erwähnt werden, dass zwei weitere Interaktionen auf dem 5%-Niveau signifikant wurden. Zunächst signalisiert hier die Interaktion A\*C (*Schulart\*Schülergeschlecht*:  $F_{(1/3201)} = 4,77$ ;  $p < .05$ ;  $Eff\% = 0,13$ ), dass der Schularteffekt bei den Jungen (*Differenz* = 0,5 *Notenstufen*) deutlicher ausgeprägt ist als bei den Mädchen (*Diff.* =

0,38). Dann weist noch die Interaktion A\*D (*Schulart\*Lehrkraftgeschlecht*:  $F_{(1/3201)} = 4,33$ ;  $p < .05$ ;  $Eff\% = 0,12$ ) darauf hin, dass die Geschlechtsunterschiede bei den Lehrkräften nur an der Hauptschule deutlicher beobachtbar sind ( $M_{\text{♂}} = 3,24$ ;  $M_{\text{♀}} = 3,12$ ), während beide Geschlechter an der Realschule gleiche durchschnittliche Zensuren erteilen ( $M = 2,74$ ).

## 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. Die Professionalisierung der Lehrerinnen dürfte große Fortschritte gemacht haben. 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 erfolgreicherer 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 (bei Carter 1952 der Mathematikleistungen) erlauben zu können, weil ihre Professionalität nicht kritisch hinterfragt wurde. Die Veränderung der Beurteilungen der Lehrerinnen hin zu einer Angleichung an die der Lehrer spräche also eher für eine Etablierung der Lehrerinnen als anerkannte pädagogische Fachkräfte. Nur eine deutlichere Bestätigung der seinerzeit 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.

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 Notenggebung der Lehrerinnen dafür zu sprechen, dass sie ihre Rolle in der Schule als eher mütterlich und damit näher am Kind verstehen. Lehrerinnen könnten sich eher in die psychische Verfassung von Schülern einfühlen, die schlechte Noten erhalten. Ob diese generelle Einstellung von allen Lehrerinnen in allen Unterrichtsfächern allerdings geteilt wird, ist bisher nicht überprüft. So könnte man sich fragen, ob das von den Lehrerinnen gewählte Unterrichtsfach nicht auch Auswirkungen auf das Urteilsverhalten haben könnte. Gerade wegen der stärker ausgeprägten sprachlichen Literalität von Mädchen und Frauen (PISA 2000 bis 2009, IGLU, LAU, DESI, VERA) wäre ja geradezu zu erwarten, dass in einem Studienfach wie dem Fach Deutsch an einer lehrerbildenden Hochschule besonders viele Frauen studieren. Zumindest für unsere Hochschule kann das bestätigt werden. Möglicherweise erwarten die so in die Schule drängenden Lehrerinnen ein ähnlich gelagertes Interesse an sprachlicher Literalität bei allen ihren Schülerinnen und Schülern und beurteilen das Bemühen darum besonders positiv. Da aber vor allem Schülerinnen sich diesen Erwartungen eher anpassen, weil es mit ihren Geschlechtsrollestereotypen kompatibel erscheint (vgl. Budde 2009), bekommen sie dann auch von den Lehrerinnen bessere Noten als die Jungen. Diese Ergebnisse dieser Untersuchung scheinen ansatzweise für diese Tendenz zu sprechen, auch wenn hier kein eindeutiger Beweis gelang.

Die Frage allerdings, ob die immer wieder postulierte mildere Beurteilungstendenz bei Lehrerinnen eine typisch weibliche Eigenschaft sei, die durchgängig in allen Fächern auftritt,

würde man nach Kenntnis der hier vorgelegten Ergebnisse wohl eher verneinen. Zunächst sei noch einmal hervorgehoben, dass die mildere Beurteilungstendenz in dieser Untersuchung zwar statistisch signifikant, aber wenig relevant ist. Hyde (2005) verweist in einer Meta-Analyse darauf, dass bei *gender*-Analysen meist große Stichproben untersucht werden, bei denen dann bereits kleine Mittelwertsunterschiede signifikant werden. Eine Überinterpretation solcher Unterschiede könne durchaus Schaden anrichten und zu kontraproduktiven politischen Entscheidungen beitragen. Unser Ergebnis kann also bei der geringen Varianzklärung nicht als Bestätigung unserer Hypothese angesehen werden. Hinzu kommt noch die Tatsache, dass bei einer gleich gelagerten Untersuchung zum Urteilsverhalten im Fach Mathematik die Beurteilungsunterschiede genau anders herum lagen (s. Birkel 2011). Dort gaben die Lehrerinnen signifikant, aber ebenso wenig relevant strengere Zensuren. Folglich spielt offenbar das untersuchte Unterrichtsfach eine Rolle, wenn man nach *gender*-Unterschieden Ausschau hält. Das würde aber gegen die generell mildere Urteilstendenz von Lehrerinnen sprechen.

Die Signifikanz des Faktors C (= Schülergeschlecht) deutet darauf hin, dass sich das Fach Deutsch erwartungsgemäß als Mädchendomäne erweist. Bei einer Varianzklärung von knapp 3% kann man den Geschlechtsunterschied durchaus auch als relevant ansehen. In den Klassen 5 und 6 schreitet aber vermutlich 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. Auch die IGLU- (Bos et al. 2007) und VERA-Ergebnisse (Isaac et al. 2006) zeigen in der Grundschule beginnende Kompetenzvorteile für die Mädchen auf. In der Sekundarstufe I scheinen sich die Kompetenzvorteile der Mädchen eher noch zu verstärken (DESI-Konsortium 2006, Deutsches PISA-Konsortium 2001, 2004, 2007, zusammenfassend vbw 2009, Klieme et al. 2010). Dass hier zu Beginn der Sekundarstufe noch Einschränkungen zu machen sind, belegen die Ergebnisse der LAU-Studie (LAU 7: Lehmann et al. 1999). In den Klassen 5 und 6 sind die mit Tests gemessenen Kompetenzunterschiede zwischen den Jungen und Mädchen zum Teil erst ansatzweise zugunsten der Mädchen erkennbar. So zeigt sich, dass zu Beginn der Klasse 5 an den Haupt- und Realschulen der Geschlechtsunterschied im Untertest „Sprache“ vernachlässigt werden kann. Erst am Ende der Klasse 6 erringen die Mädchen hier einen deutlicheren Vorteil ( $d=.01 \rightarrow d=.19$ ), weil die Jungen geringere Lernzuwächse zeigen als die Mädchen. Beim Untertest „Leseverständnis“ zeigt sich bereits zu Anfang der Klasse 5 ein Vorsprung der Mädchen, der sich aber bis zum Ende der Klasse 6 kaum verändert ( $d=.10 \rightarrow d=.13$ ). Nur beim Testteil „Rechtschreibung“ zeigt sich eine recht deutliche Überlegenheit der Mädchen bereits am Anfang der Klasse 5, die auch bis zum Ende der 6. Klasse aufrecht erhalten werden kann ( $d=.39 \rightarrow d=.41$ ). Bei den Jungen wurden z.T. extrem schwache Rechtschreibleistungen diagnostiziert. Lehrkräfte, die im Fach Deutsch nun eine Gesamtnote erteilen müssen, werden mit individuell unterschiedlicher Gewichtung nicht nur diese Teilkompetenzen in ihre Zensur eingehen lassen, sondern auch noch weitere Leistungen wie z.B. die mündliche Beteiligung am Unterricht, die Zuverlässigkeit der Hausaufgabenerledigung und eventuell noch weitere Aspekte. Bedenkt man dann noch, dass in den zurück liegenden Jahren die ehemals hohe Bedeutung der Rechtschreibung für die Deutschnoten zu reduzieren versucht wurde, dann dürfte klar werden, dass die zwar hoch signifikanten, mit 3% Varianzklärung aber doch eher am unteren Bereich der Relevanz sich bewegenden Notenvorteile der Mädchen in dieser Untersuchung zu verstehen sind. Sie honorieren vielleicht auch das schulangepasste Verhalten der Mädchen (s. Spinath et al. 2010), das sich in einer unterrichtsförderlichen Weise in der mündlichen Beteiligung und der höheren Zuverlässigkeit bei Aufgabenerledigungen niederschlägt.

Die besseren Zensuren der Mädchen im Fach Deutsch in dieser Untersuchung stehen also im Einklang mit der berichteten Entwicklung der Kompetenzvorteile von Mädchen in den sprachlichen Fächern. Zu bedenken ist allerdings an dieser Stelle, dass die signifikanten Mittelwertsunterschiede trotz allem nicht bedeuten, dass nicht auch Jungen gute und sehr gute No-

ten erhalten können. In Abbildung 1 wird sichtbar, dass sich die Streubereiche bei beiden Geschlechtern über die gesamte Notenskala verteilen. Nur die Verteilungsschwerpunkte unterscheiden sich. Vielleicht hätte man sogar noch größere Beurteilungsunterschiede im Fach Deutsch zugunsten der Mädchen erwarten können, aber möglicherweise schlagen die besonders deutlich erst im weiteren Verlauf der Beschulung durch, wenn in der Pubertät das Verhalten der Jungen im Deutschunterricht noch weniger unterrichtsförderlich wird<sup>4</sup> (Budde (2009, S. 35, Spinath et al. 2010). Lehrkräfte könnten dann das schulunangepasste Verhalten der Jungen noch stärker durch schlechtere Zensuren abstrafen oder auf die schlechter werdenden Leistungen der Jungen reagieren.

Die Entwicklung sprachlicher Kompetenzen wird von Mädchen offenbar als deutlich stärker erstrebenswert und mit dem *gender*-Stereotyp kompatibel (Steffens et al. 2010) erlebt und deshalb im weiteren Verlauf der Schulzeit immer stärker ausgeprägt als bei den Jungen (Nyssen 2004, Kampshoff 2007). Implizite *gender*-Stereotype bezüglich der Bedeutung von im Deutschunterricht dokumentierten Kompetenzen beeinflussen das Selbstkonzept und die schulischen Leistungen von Mädchen bereits in der Grundschule und gewinnen im Verlauf der Entwicklung in der Sekundarstufe I weiter an Bedeutung (Steffens et al. 2010). Solche Stereotype werden u.U. von den Jungen sogar als bedrohlich erlebt. Jungen halten unterrichtsstörendes Verhalten und geringere Kompetenzen im sprachlichen Bereich als eher kompatibel mit ihrem *gender*-Stereotyp. Dass sie sich damit Nachteile bei der Leistungsbeurteilung im Fach Deutsch einhandeln, nehmen sie eher in Kauf als ein Verhalten, das sich gegen das *gender*-Stereotyp interpretieren lässt (zum „*stereotype threat*“ siehe Good 2001).

Wenn die Beurteilungsunterschiede zwischen den Mädchen und Jungen hier nicht ganz so deutlich auftreten, wie man es vielleicht erwartet hätte, dann sind hier u.U. bereits erste Erfolge zu erkennen, die darauf abzielen, den Unterricht *gender*-sensibel so zu gestalten (Budde, Scholand & Faulstich-Wieland 2008), dass ein Erfolg der Jungen wahrscheinlicher wird. Nachdem die Unterrichtsinhalte in den vergangenen Jahrzehnten immer deutlicher auf die Interessen der Mädchen umgestellt worden zu sein scheint, könnte die Berücksichtigung jugendspezifischer Themen jetzt dazu beigetragen haben, deren Bereitschaft, sich auf unterrichtliche Themen im Fach Deutsch wieder stärker einzulassen, zu unterstützen. Guiso et al. (2008) weisen zudem darauf hin, dass in Ländern mit einer relativ ausgeprägten Chancengleichheit für beide Geschlechter sich die Geschlechtsunterschiede bei den Kompetenzen im Fach Mathematik deutlich verringern. Generell korreliert z.B. der Index für weibliche Emanzipation (GGI) in den Teilnehmerländern der PISA-Studien zu  $-0,57$  mit der Stärke der männlichen Dominanz. Vielleicht lassen sich solche Ergebnisse ja auch anders herum interpretieren als eine Tendenz zur Angleichung der Beurteilungen im Fach Deutsch, wenn es tatsächlich gelänge, auch hier eine größere *gender*-Sensibilität im Unterricht zu realisieren. Neben der *gender*-Sensibilität des Unterrichts sind aber auch weitere Faktoren denkbar, die für die Überlegenheit der Mädchen im Fach Deutsch zumindest mitverantwortlich gemacht werden können. Eine Auflistung möglicher Faktoren findet sich bei Hannover (2010 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 Leistun-

---

<sup>4</sup> Schulunangepasstes Verhalten führt bei Jungen und bei Mädchen zu einer negativeren Beurteilung. Dieses Verhalten kommt nur bei Jungen viel häufiger vor.



gen war. Für die Vorhersage des allgemeinen Schulerfolgs (GPA) erwies sich zudem das Leistungsselbstkonzept 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 die einzelnen Schulfächer (Spinath et al. 2010), so erwiesen sich sowohl die „Big Five“-Persönlichkeitsfaktoren als auch die 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 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.

Auch wenn die Interaktion zwischen dem Geschlecht der Lehrkräfte und dem der Schülerinnen und Schüler nicht besonders relevant war, und man sie sicher nicht überinterpretieren darf, was in Übereinstimmung zu sehen ist mit den Ergebnissen von Marsh et al. (2008), ist ihre Signifikanz an dieser Stelle doch eine Betrachtung wert, weil sie in Übereinstimmung steht mit den Untersuchungsergebnissen von Carter (1952). Zunächst einmal fällt auf, dass Lehrerinnen den Jungen im Schnitt genau die gleiche Zensur erteilten wie die Lehrer. Bei der Beurteilung der Jungen spielt also das Geschlecht der Lehrkraft überhaupt keine Rolle. Dann ist da die Tatsache, dass die Lehrerinnen den Mädchen einen Bonus von 0,14 Notenstufen gegenüber den Lehrern geben. Sie honorieren damit möglicherweise deren ausgeprägtere Literalität die ihnen als Lehrerinnen besonders an Herzen liegt. Sie selbst haben ja bereits während der eigenen Schulzeit und im beruflichen Werdegang deren überaus hohe Bedeutung kennengelernt. Nun setzen sie diese bei ihren Schülerinnen und Schülern als ebenso gegeben voraus und erwarten eine ähnlich gelagerte Begeisterung und ein entsprechendes Unterrichtsverhalten im Fach Deutsch. Solche meist impliziten Normvorstellungen werden aber vor allem von den Mädchen akzeptiert und internalisiert, weil sie mit dem weiblichen Rollenstereotyp kompatibel erscheinen (Bacher et al. 2008). Jungen tun sich schwer, weil ihre Interessen anders gelagert sind und die Literalität als dem männlichen Rollenstereotyp abträglich erscheint. Ihr externalisierendes Verhalten im Fach Deutsch (Spinath et al. 2010) stört den harmonischen Fortgang des Unterrichts. Das aber ist für Lehrerinnen inkompatibel mit der impliziten Vorstellung bezüglich der Literalität von Schülerinnen und Schülern. An dieser Stelle muss aber noch einmal auf die geringe Relevanz trotz statistischer Signifikanz hingewiesen werden. Solche Interpretationsversuche sind also mit größter Vorsicht zu behandeln!

Eine andere Erklärung wäre aber – bei gleicher Vorsicht der Interpretation – ebenso denkbar. Alle nationalen, wie auch internationalen Vergleichsuntersuchungen zeigen den großen Kompetenzvorsprung der Mädchen bei Leistungen, die dem Unterrichtsfach Deutsch zuzurechnen sind. Wenn diese Leistungen eine wesentliche Grundlage für die Notengebung in den Schulen spielen, dann müssten alle Lehrkräfte diese Vorsprünge angemessen in Ihre Zeugniszensuren eingehen lassen. Die Frage wäre dann nur, ob die bessere Beurteilungen der Mädchen gegenüber den Jungen durch die Lehrer als ausreichend anzusehen ist. Vielleicht bildet die noch bessere Beurteilung durch die Lehrerinnen diesen Vorsprung ja sogar besser ab als der, den die Lehrer zu geben bereit sind. Eine „systematische Ungleichbehandlung der Jungen und Mädchen“ (Valtin et al. 2005) daraus ableiten zu wollen, erscheint als sehr gewagt! Eine Entscheidung dieser Frage ist bei der augenblicklichen Datenlage nicht möglich, könnte aber vielleicht einmal genauer untersucht werden.

## **6. Literatur**

Arnot, M.; David, M. & Weiner, G. (1998). *Closing the gender gap*. Polity Press: Cambridge.

- Bacher, J.; Beham, M. & Lachmayr, N. (2008). Geschlechterunterschiede bei der Bildungswahl. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Balluseck, H. von (2009). Jungen als Verlierer? Die Diskussion um die Benachteiligung von Jungen in der Schule. *querelles-net*, 10, Nr.1. aufgerufen am 4.11.2009 unter:  
<http://www.querelles-net.de/index.php/qn/article/viewArticle/726/734>
- Baumert, J. et al. (1997). *TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde*. Leske & Budrich: Opladen.
- 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*. Leske & Budrich: Opladen.
- Birkel, P. (1978). *Mündliche Prüfungen. Zur Objektivität und Validität der Leistungsbeurteilung*. Kamp: Bochum.
- 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). Geschlechtsspezifische Effekte bei der Zensurenggebung im Fach Mathematik? *Verfügbar unter: http://www.*
- Bos, W.; Hornberg, S.; Arnold, K.H.; Faust, G.; Fried, L.; Lankes, E.M.; Schwippert, K. & Valtin, R. (Hrsg.). (2007). IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich. Münster: Waxmann.
- Budde, J. (2008). *Bildungs(miss)erfolge von Jungen und Berufswahlverhalten bei Jungen/ männlichen Jugendlichen*. Berlin: BMBF, Referat Chancengleichheit in Bildung und Forschung.
- Budde, J. (2009). *Mathematikunterricht und Geschlecht. Empirische Ergebnisse und pädagogische Ansätze*. Bildungsforschung Band 30, herausgeg. vom Bundesministerium für Bildung und Forschung. Bonn, Berlin.
- 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.
- 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.) (1995<sup>9</sup>). *Die Fragwürdigkeit der Zensurenggebung*. S. 148-158. Weinheim: Beltz.
- Day, L.C. (1938). Boys and girls and honour ranks. *School Review*, 46, S. 288-299.
- Dee, T.S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources*, 42, S. 528 – 554.
- DESI-Konsortium (Hrsg.). (2006). Unterricht und Kompetenzerwerb in Deutsch und Englisch. Zentrale Befunde der Studie Deutsch-Englisch-Schülerleistungen-International. Frankfurt/M.: DIPF.
- 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. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Münster: Waxmann.
- Deutsches PISA-Konsortium (Hrsg.). (2007). *PISA 2006. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Münster: Waxmann.
- Diefenbach, H. (2008). Jungen und schulische Bildung. In: M. Matzner & W. Tischner (Hrsg.). *Handbuch Jungen-Pädagogik*. (S. 92 – 108). Weinheim: Beltz.
- Diefenbach, H. & Klein, M. (2002). „Bringing Boys Back in“. Soziale Ungleichheit zwischen den Geschlechtern im Bildungssystem zuungunsten von Jungen am Beispiel der Sekundarschulabschlüsse. *Zeitschrift für Pädagogik*, 48, S. 938-958.

- Douglass, H.R. (1937). Relation of high-school marks to sex in four Minnesota senior high schools. *School Review*, 45, S. 282-288.
- Driessen, G. (2007). The feminization of primary education: Effects of teachers' sex on pupil achievement, attitudes and behaviour. *International Review of Education*, 53, S. 183 – 203.
- Edminston, R.W. (1943). Do teachers show partiality toward boys or girls? *Peabody Journal of Education*, 20, S. 234-238.
- Freudenthaler, H.; Spinath, B. & Neubauer, A. (2008). Predicting school achievement in boys and girls. *European Journal of Personality*, 22, S. 231-245.
- Garbe, Ch. (2007). Lesen – Sozialisation – Geschlecht. Geschlechterdifferenzierende Leseforschung und –förderung. In: A. Bertschi-Kaufmann (Hrsg.). *Lesekompetenz – Leseleistung – Leseförderung. Grundlagen, Modelle und Materialien*. Seelze: Friedrich Kallmeyer.
- Garbe, Ch. (2008). „Echte Kerle lesen nicht!“ – Was eine erfolgreiche Leseförderung für Jungen beachten muss. In: M. Matzner & W. Tischner (Hrsg.). *Handbuch Jungen-Pädagogik*. S. 301-315. Weinheim: Beltz.
- Garbe, Ch. (2011). „Richtige Jungs“ lesen nicht? *Erziehung und Wissenschaft. Zeitschrift der Bildungsgewerkschaft GEW*, 63, Heft 3, S. 26+28.
- Giest, H. (1996). Besonderheiten des Problemlösens bei Schülern 4., 6. und 8. Klassen. *Empirische Pädagogik*, 310, 3, S. 307-342.
- Helbig, M. (2010). Sind Lehrerinnen für den geringen Schulerfolg von Jungen verantwortlich? *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 62, S. 93 – 111.
- Heller, K. & Rosemann, B. (1974). *Planung und Auswertung empirischer Untersuchungen*. Stuttgart: Klett.
- Horstkemper, M. (1999). Ausgewählte Aspekte der Lehrer-Schüler-Interaktion. In: M. Horstkemper (Hrsg.). *Koedukation. Erbe und Chancen*. (S. 139 – 158). Weinheim: Juventa.
- Hyde, J.S. (2005). The gender similarities hypothesis. *American Psychologist*, 60, S. 581-592.
- Ingenkamp, K. (Hrsg.) (1971, 1995). *Die Fragwürdigkeit der Zensurengebung*. Weinheim: Beltz. 9. Auflage 1995.
- Ingenkamp, K. & Lissmann, U. (2005). *Lehrbuch der Pädagogischen Diagnostik*. Weinheim: Beltz.
- Isaac, K.; Halt, A.C.; Hosenfeld, I.; Helmke, A. & Groß-Ophoff, J. (2006). VERA. Qualitätsentwicklung und Lehrerprofessionalisierung durch Vergleichsarbeiten. *Deutsche Schule*, 98, 1, S. 107-110.
- Kampshoff, M. (2007). *Geschlechterdifferenz und Schulleistung*. Wiesbaden: Verlag für Sozialwissenschaften.
- Kleiter, E. (2004). *KMSS-8. Kleiter-Microcomputer-Statistik-System*. Kiel
- Klein, J. (2004). Who is most responsible for gender differences in scholastic achievement: pupils or teachers? *Educational Research*, 49, Heft 2, S. 183-193.
- Klieme, E.; Artelt, C.; Hartig, J.; Jude, N.; Köller, O.; Prenzel, M.; Schneider, W. & Stanat, P. (Hrsg.) (2010). *PISA 2009. Bilanz nach einem Jahrzehnt*. Münster: Waxmann.
- Kuhn, H.P. (2008). Geschlechterverhältnisse in der Schule. In: Rendtorff, B. & Prengel, A. *Kinder und ihr Geschlecht*. S. 49-71. Opladen: Budrich.
- Lehmann, R.H.; Gänsfuß, R. & Peek, R. (1999). Aspekte der Lernausgangslage und der Lernentwicklung von Schülerinnen und Schülern, die im Schuljahr 1996/97 eine fünfte Klasse an Hamburger Schulen besuchten. Bericht über die Erhebung im September 1998 (LAU 7). Aufgerufen am 18.08.2010 unter: [www.hamburger-bildungsserver.de/schulentwicklung/lau/lau7/Litverz.htm](http://www.hamburger-bildungsserver.de/schulentwicklung/lau/lau7/Litverz.htm)
- 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.
- Martino, W.J. (2008). Male teacher as role models: Addressing issues of masculinity, pedagogy and re-masculinisation of schooling. *Inquiry*, 38, S. 189 – 223.
- 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.
- OECD (2010). PISA 2009 at a Glance. Aufgerufen am 3.2.2011 auf : [www.oecd.org/edu/pisa/2009](http://www.oecd.org/edu/pisa/2009).
- Peek, R. (2008). Kompetenzen und Kompetenzmessung im Kontext von Fachdidaktik, Psychometrie und Unterrichtsentwicklung. In: M. Böhnisch (Hrsg.). *Beiträge zum 16. Symposium Deutschdidaktik "Kompeten-*

- zen im Deutschunterricht". Sonderheft: Didaktik Deutsch. S. 162-172. Baltmannsweiler: Schneider Verlag Hochgehren.
- Philipp, M. & Garbe, Ch. (2007). Lesen und Geschlecht – empirisch beobachtbare Achsen der Differenz. In: A. Bertschi-Kaufmann (Hrsg.). Lesekompetenz – Leseleistung – Leseförderung. Grundlagen, Modelle und Materialien. Seelze: Freidrich Kallmeyer.
- Preuss-Lausitz, U. (2005). Anforderungen an eine jungenfreundliche Schule. *Die Deutsche Schule*, 97, S.222 – 235.
- Rohrmann, T. (2008). Jungen in der Grundschule. In: M. Matzner & W. Teschner (Hrsg.). *Handbuch der Jungen-Pädagogik*. S. 109-121. Weinheim: Beltz.
- Sexton, P. (1969). *The feminized male: classroom white collars, and the decline of manliness*. New York: Random House.
- 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.
- Rose, L. & Schmauch, U. (2005). *Jungen – die neuen Verlierer?* Königstein: Helmer.
- 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/](http://www.erzwiss.uni-hamburg.de/degendering_science/) (11.11.2009)
- Schrader, F.-W. & Helmke, A. (2001). Alltägliche Leistungsbeurteilung durch Lehrer. In F.E. Weinert (Hrsg.) *Leistungsmessungen in Schulen*. (S. 45-58). Weinheim: Beltz.
- Schultheis, K. (2008). Jungenforschung. Aktuelle Ergebnisse, Desiderate, Probleme. In: M. Matzner & W. Teschner (Hrsg.). *Handbuch der Jungen-Pädagogik*. S. 63-77. Weinheim: Beltz.
- Spinath, B.; Freudenthaler, H. & Neubauer, A. (2010). Predicting domain-specific school achievement in boys and girls by intelligence, personality and motivation. *Personality and Individual Differences*, 48, S. 481-486.
- Stamm, M. (2008). Underachievement von Jungen: Perspektiven eines internationalen Diskurses. *Zeitschrift für Erziehungswissenschaft*, 11, S. 106 – 124.
- Stürzer, M. (2003a). Geschlechtsspezifische Schulleistungen. In: M. Stürzer, R. Roisch, A. Hunze & W. Cornelißen. *Geschlechterverhältnisse in der Schule*. S. 83-116. Opladen: Leske & Budrich.
- Stürzer, M. (2003b). Unterrichtsformen und die Interaktion der Geschlechter in der Schule. In: M. Stürzer, R. Roisch, A. Hunze & W. Cornelißen. *Geschlechterverhältnisse in der Schule*. S. 151ff. Opladen: Leske & Budrich.
- vbw – Vereinigung der Bayerischen Wirtschaft e.V. (Hrsg.). (2009). *Geschlechterdifferenzen im Bildungssystem. Jahresgutachten 2009*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Wallner-Paschon, Ch. (2010). *Notengerechtigkeit bei der Risiko- und Spitzengruppe | bifie*. Aufgerufen am 24.11.2010 unter: [www.bifie.at/buch/395/4/3](http://www.bifie.at/buch/395/4/3)
- [www.frauensicht.ch/aktuell/](http://www.frauensicht.ch/aktuell/) (aufgerufen am 13.12.2004)
- [www.linzerin.at/themen/artikel\\_drucken.php?id=44](http://www.linzerin.at/themen/artikel_drucken.php?id=44) (aufgerufen am 13.12.2004)

---

Autor: **Dr. Peter Birkel**, PH Weingarten,  
Doggenriedstr. 18,  
88250 Weingarten  
E-Mail: [birkel@ph-weingarten.de](mailto:birkel@ph-weingarten.de)

Der **Datensatz** dieser Untersuchung wird **auf Anfrage** gern zur weiteren Bearbeitung **zur Verfügung gestellt**.