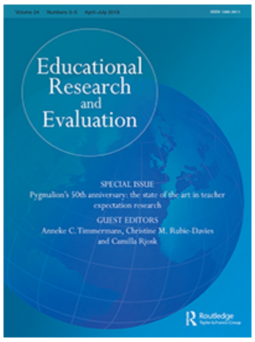

Routledge
Taylor & Francis Group

Educational Research and Evaluation

An International Journal on Theory and Practice

ISSN: 1380-3611 (Print) 1744-4187 (Online) Journal homepage: https://www.tandfonline.com/loi/nere20

Pygmalion and the gender gap: do teacher expectations contribute to differences in achievement between boys and girls at the beginning of schooling?

Sarah Gentrup & Camilla Rjosk

To cite this article: Sarah Gentrup & Camilla Rjosk (2018) Pygmalion and the gender gap: do teacher expectations contribute to differences in achievement between boys and girls at the beginning of schooling?, Educational Research and Evaluation, 24:3-5, 295-323, DOI: 10.1080/13803611.2018.1550840

To link to this article: https://doi.org/10.1080/13803611.2018.1550840

Published online: 12 Dec 2018.

EDUCATIONAL RESEARCH AND EVALUATION
2018, VOL. 24, NOS. 3–5, 295–323
https://doi.org/10.1080/13803611.2018.1550840

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Pygmalion and the gender gap: do teacher expectations contribute to differences in achievement between boys and girls at the beginning of schooling?

Sarah Gentrup[a] and Camilla Rjosk[b]

[a]Institute for Educational Science, Humboldt-Universität zu Berlin, Berlin, Germany; [b]Institute for Educational Quality Improvement, Humboldt-Universität zu Berlin, Berlin, Germany

ABSTRACT
This study examined the role of teacher expectations in the emerging gender gaps in reading and mathematics in the first year of schooling. Therefore, we first investigated whether boys and girls differ in their vulnerability to teacher expectancy effects. Second, we analysed whether gender-specific effects of teacher expectations contribute to gender achievement gaps. Our analyses were based on 1,025 first-grade students in Germany. Among the majority of the students, boys and girls did not differ in their vulnerability to teacher expectancy effects. Further analyses examined a subgroup of students who were targets of relatively strong teacher expectation bias and who showed unexpectedly high or low achievement gains. In this specific subgroup, girls' mathematics achievement was more adversely affected by negatively biased expectations and benefitted less from positive bias than boys' achievement. Mediation analyses revealed that teacher expectation bias did not substantially contribute to gender gaps in reading or mathematics.

KEYWORDS
Educational disparities; gender gap; mathematics; reading; self-fulfilling prophecy; teacher expectations

Introduction

The issue of gender disparities in students' academic achievement has a long tradition in educational research. In almost all Western countries, significant gender differences in students' proficiency in reading and, to a lesser extent, in mathematics occur in the first years of schooling. Large-scale assessment studies in schools, such as the Progress in International Literacy Study (PIRLS), the Trends in International Mathematics and Science Study (TIMSS), and the Programme for International Student Assessment (PISA), give us information on gender disparities at different age levels. In all of these studies, test scores are scaled to an overall mean of 500 and a standard deviation of 100. As indicated by gender-specific analyses, fourth-grade girls outperform boys in reading by an average of 13 points across Organisation for Economic Co-operation and Development (OECD) countries (PIRLS: McElvany, Kessels, Schwabe, & Kasper, 2017; Mullis, Martin, Foy, & Hooper, 2017, p. 36). In mathematics, a small

CONTACT Sarah Gentrup sarah.gentrup@hu-berlin.de

advantage of 5 points for boys has been observed (TIMSS: Mullis, Martin, Foy, & Hooper, 2016; Wendt, Steinmayr, & Kasper, 2016). The gender gaps in both subjects further increase at the lower secondary educational levels. In the PISA 2015 test, 15-year-old girls outperformed their male peers in reading achievement by 27 points on average across OECD countries, whereas boys scored an average of eight points higher on mathematics tests than girls (OECD, 2016, pp. 168–169, 196–197). Although these gender gaps in reading and, to a lesser degree, mathematics in the later school years are substantial, boys' and girls' linguistic and mathematical skills are more similar upon school entry (Fryer & Levitt, 2010; Mengering, 2005; Robinson & Lubienski, 2011). This evidence raises the question of why gender gaps emerge and widen during the early years of schooling. According to social-psychological theories of socialisation, gender differences in academic interests and behaviours are likely due in part to the influences of significant others, such as parents, peers, and teachers. These theories assume that significant others confirm or disconfirm a child's behaviour based on their gender-specific expectations towards that child (cf. Trautner, 2008). With school enrolment, teachers play an important role in students' reading acquisition and mathematics learning. Thus, their gender-specific expectations may be important facilitators in the widening of gender gaps. The current study investigated this assumption and examined whether the effects of gender-specific teacher expectations of students' performance contributed to differences in reading and mathematics achievement between boys and girls at the beginning of schooling. The analyses were based on data from a study specifically designed to investigate teacher expectancy effects among first-grade students in Germany.

Teacher expectations towards boys and girls

Many studies indicate that teacher expectations of student performance differ by student gender, regardless of their actual achievement (gender bias). Nevertheless, the evidence remains somewhat inconsistent. Some international studies have indicated that teachers overestimate the linguistic skills of girls more often than those of boys (Hinnant, O'Brien, & Ghazarian, 2009; Ready & Wright, 2011), whereas boys have been more favourably regarded in mathematics (Jussim & Eccles, 1992), even when girls and boys show similar achievement levels. However, some studies have not found any differences in teacher expectations for boys and girls when controlling for students' actual achievement (McKown & Weinstein, 2008; Van den Bergh, Denessen, Hornstra, Voeten, & Holland, 2010). In Germany, the situation is similar. Some studies have found no gender bias in teacher perceptions regarding student achievement in reading (Karing, Matthäi, & Artelt, 2011) or mathematics (Schrader & Helmke, 1990). Other studies have identified that teachers consider male students to perform better in mathematics than female students when actual student achievement is kept constant (Holder & Kessels, 2017; Tiedemann, 2002). Stahl (2007) reported overestimations of girls' performance in reading. Similarly, the study by Lorenz, Gentrup, Kristen, Stanat, and Kogan (2016) indicated a positive bias of teacher expectations in linguistics for girls compared to boys. No gender bias in mathematics was identified in this study.

However, these studies have investigated gender differences in teacher expectations when accounting for students' actual achievement but not when accounting for teacher perceptions of further student learning resources. As argued by Schrader and Helmke (1990), as well as Robinson-Cimpian, Lubienski, Ganley, and Copur-Gencturk (2014), teacher expectations are likely to be coloured by teacher perceptions of student characteristics other than achievement. Substantial overlaps for different indicators of teacher expectations of student achievement and teacher perceptions of student motivation and work habits have been identified. For example, a student seen by his or her teacher as motivated and hardworking is more likely to be considered more talented and better performing (Anders, McElvany, & Baumert, 2010; Schrader & Helmke, 1990) and more likely to receive better grades (Anders et al., 2010; Kuhl & Hannover, 2012) and track recommendations (Anders et al., 2010; Timmermans, De Boer, & Van der Werf, 2016) than their lower estimated classmates who, in fact, perform similarly on achievement tests (cf. Gentrup, Rjosk, Stanat, & Lorenz, 2018).

As teachers generally perceive girls as being more motivated and more eager to learn (e.g., Anders et al., 2010; Jussim & Eccles, 1992; Jussim, Eccles, & Madon, 1996) and as behaving better at school (e.g., Glock, 2016; Glock & Kleen, 2017), the gender differences in teacher expectations may be partly attributable to teacher perceptions of student motivation and work habits. Following this assumption, Robinson-Cimpian et al. (2014) analysed gender bias in teacher perceptions of students' mathematics achievement when their perceptions included their ratings of students' behaviour and effort. Whereas no gender differences were found in teacher perceptions of students' mathematics performance when controlling for students' current achievement, an advantage for boys became apparent when the authors also took into account teacher perceptions of student behaviour and effort (Robinson-Cimpian et al., 2014). Similar results were found by Gentrup, Rjosk, et al. (2018), indicating that teachers expected girls to perform equivalently to boys in mathematics only if the teachers observed girls as being more eager to learn and as working harder than their male classmates. In contrast, the female advantage in teacher expectations regarding student reading performance could be fully explained by teacher perceptions of student motivation and work habits (Gentrup, Rjosk, et al., 2018). Robinson-Cimpian et al. (2014) referred to this phenomenon as a conditional over- and underestimation. As an illustrative example, take the result that student gender did not significantly predict teacher expectations in mathematics when accounting for actual achievement. If this result indicated unbiased expectations, no gender difference should occur when boys and girls share similar levels of other characteristics, such as teacher perceptions of student motivation and work habits. The more similar boys and girls were in terms of various characteristics, the smaller the gender differences in teacher expectations should become (cf. Robinson-Cimpian et al., 2014). If taking teacher perceptions of student motivation and work habits into account increases gender gaps, we refer to this situation as a conditional over- or underestimation.

In summarising the results based on this logic, it seems reasonable to assume that gender-specific teacher expectations exist but depend to some degree on teacher perceptions of student motivation and work habits. Such gender-specific teacher expectations may contribute to gender gaps through the process of self-fulfilling teacher expectations or teacher expectancy effects (Merton, 1948; Rosenthal & Jacobson, 1968).

Teacher expectancy effects

The research on the effects of teacher expectations traces back to the experiment "Pygmalion in the Classroom" by Rosenthal and Jacobson (1968), testing the idea of self-fulfilling prophecies in the context of a school. It is widely assumed that self-fulfilling prophecies in the classroom follow a sequence of three major steps (Jussim, Robustelli, & Cain, 2009): (1) teachers form inaccurate expectations, that is, expectations that are too high or too low compared to actual student characteristics, such as achievement and motivation; (2) teachers treat high- and low-expectancy students differently; and (3) students react to this differential treatment in a way that confirms the teachers' initial expectations. The assumption that teacher expectancy effects emerge and, thus, that teacher expectations may influence students' learning in the direction of their expectations is widely supported (for overviews, see Jussim et al., 1996; Jussim et al., 2009). Most studies stem from the US, but there are some studies from Europe as well (e.g., Germany: Friedrich, Flunger, Nagengast, Jonkmann, & Trautwein, 2015; Gentrup, Lorenz, Kristen, & Kogan, 2018; Lorenz, 2018; Stahl, 2007; The Netherlands: De Boer, Bosker, & Van der Werf, 2010; Jungbluth, 1993). However, teacher expectancy effects are generally small in magnitude. Measured as effect size d, the average effect was $d = .30$ (Jussim et al., 2009). When investigating gender disparities in student achievement, there are two possible mechanisms through which teacher expectancy effects may contribute to gender gaps. The two mechanisms should not be viewed as mutually exclusive; rather, they may complement each other.

First, boys and girls may differ in their vulnerability to teachers' biased expectations (moderation perspective). Specifically, boys in reading and girls in mathematics may be more vulnerable to teachers' underestimations, as the lower expectations would be compatible with common gender stereotypes. For the same reason, girls in reading and boys in mathematics may benefit more strongly from teachers' overestimations. This mechanism may contribute to gender disparities even if teacher expectations are not systematically biased by student gender (but are randomly inaccurate for individual students independent of their gender, as inaccuracy is a precondition of teacher expectancy effects). For example, a teacher may hold positively biased expectations for all of his or her students in reading, but girls may benefit more strongly from these overestimations because the high expectation is compatible with the stereotype "girls read well". As a result, girls would develop especially advanced reading skills, whereas boys' learning would increase only slightly. This process, therefore, would contribute to more advanced reading skills among girls than among boys.

Second, teacher expectations may influence student achievement through self-fulfilling prophecies, and because boys and girls have different probabilities of being over- or underestimated in reading and mathematics, teacher expectations may contribute to students' differing achievement gains (mediation perspective). In contrast to the first mechanism, this mechanism requires teacher expectations that are not only inaccurate but also systematically biased by student gender (cf. Lorenz et al., 2016). To make this precondition clear, let us assume a hypothetical class with 20 students – 10 boys and 10 girls. If boys and girls are equally likely to be overestimated by teachers, then, for example, two boys and two girls would be seen as performing better than they actually are. As a consequence of teacher expectancy effects, both boys and both girls would perform better throughout the school year, and therefore, no contribution to gender differences would emerge. Only if boys and girls are unequally likely to be over- or underestimated would such processes contribute to gender gaps; that is, if, in our hypothetical class, four boys and only two girls became targets of overestimations, more boys than girls would benefit from positive teacher expectancy effects and, as a result, would lift the average male performance above the average female performance. The state of evidence regarding both possible mechanisms will be described next.

Gender-specific teacher expectancy effects: a moderation perspective

The assumption that teacher expectancy effects might be gender specific is based on research on stereotype threat. In performance situations when stereotypes are salient, members of groups stigmatised by a stereotype may show poorer performance because of the anxiety of confirming the stereotype, which is known as stereotype threat (Steele & Aronson, 1995). Common gender stereotypes include girls being less talented in mathematics and boys being less successful in the linguistic domain, such as in reading (cf. Heyder & Kessels, 2013; Steffens & Jelenec, 2011). Thus, the negative effects of stereotypes are reasonable to assume for girls in mathematics and for boys in reading. A large number of studies support the existence of stereotype threat effects for girls in mathematics (for a meta-analysis on school-aged girls, see Flore & Wicherts, 2015). For boys in reading, stereotype threat is less often investigated, but the existing literature supports this phenomenon as well (e.g., Latsch & Hannover, 2014; Pansu et al., 2016). However, studies have observed not only negative effects of stereotypes for the stigmatised group (stereotype threat) but also positive effects for the non-stigmatised group (stereotype lift; e.g., Latsch & Hannover, 2014). In this case, the benefit in performance is triggered by a negative outgroup stereotype. This downward comparison is assumed to enhance confidence, motivation, and self-efficacy in members of the non-stigmatised group and thus result in better performance (cf. Walton & Cohen, 2003).

Following these results of stereotype threat research, boys in reading and girls in mathematics may be more vulnerable to underestimations by their teachers, as lower expectations would be compatible with common gender stereotypes. In addition to general teacher expectancy effects, these stereotype-compatible teacher expectations may increase boys' and girls' fears of confirming the gender stereotype while reducing cognitive capacity and motivation. Eventually, these expectations may lead to lower achievement of boys in reading and of girls in mathematics and thus result in achievement levels that confirm both teacher expectations and gender stereotypes. In addition, similar to stereotype lift, positive expectancy effects might be stronger for girls in reading and for boys in mathematics. Ultimately, both processes would lead to student achievement that confirms common gender stereotypes. Thus, boys would perform worse in reading, and girls would lag behind in mathematics.

To date, three studies have examined the issue of gender specificity in teacher expectancy effects; in other words, these three studies have tested the assumption of student gender serving as a moderator. Jussim et al. (1996) evaluated possible overall gender differences in the effects of teacher perceptions on mathematics test scores and final sixth-grade marks without considering gender-specific responses to negative and positive bias. Gender differences occurred only when predicting students' final sixth-grade marks in mathematics based on teacher perceptions of students' maths talent. These effects were found to be stronger for girls than for boys. In fact, teacher perceptions of maths talent were positively related to students' final marks only for girls. No gender differences were found for teacher perceptions of student performance or teacher perceptions of student effort. Furthermore, no gender differences were apparent for any of the variables when students' mathematics test performance was predicted instead of final sixth-grade marks. However, as teacher perceptions were captured as a linear predictor without including quadratic terms, the analyses could detect overall gender differences in the magnitude of the effects only. To examine the assumed reversed responses of boys and girls to positive and negative bias, a quadratic term for teacher perceptions would be necessary.

Two more recent studies took this issue into account. De Boer et al. (2010) examined moderation in the prediction of an indicator of overall school success (a score on the education ladder that combined the grade level and school track), controlling for an overall student performance score covering linguistic and mathematical achievement as well as meta-cognitive skills. Teacher expectation bias was subdivided into seven categories ranging from severe negative expectation bias to severe positive expectation bias. The category "no bias" served as the reference category. The analyses revealed no gender specificity in teacher expectancy effects for any of the six categories of teacher expectation bias. However, as an overall student performance score was investigated, the analyses allowed for no conclusion to be drawn on domain-specific moderation. Furthermore, the non-significant findings may be a result of opposing response patterns in the linguistic and mathematical domains, as one would expect based on evidence from stereotype threat research.

The study by McKown and Weinstein (2002) took this issue into account and investigated whether male and female first-, third-, and fifth-grade students differed in their response to teacher over- and underestimation in mathematics and reading, separately. After investigating linear and quadratic interactions between biased teacher expectations and student gender, which did not show any significant moderation, McKown and Weinstein (2002) divided the students into different groups. Teacher expectation bias was categorised into underestimated expectations, overestimated expectations and accurate expectations. Furthermore, student achievement development was subdivided into students whose skills improved, declined, or remained stable. Combining these two variables, students were categorised into confirmers (whose actual achievement development corresponded with the direction of the over- or underestimation by their teacher) and disconfirmers (whose actual achievement development was in opposition to their teachers' over- or underestimation). Students who were accurately estimated by their teachers and students whose skills remained stable were excluded from this categorisation and from further analyses. Comparing the resulting eight groups of male and female underestimated confirmers, overestimated confirmers, underestimated disconfirmers, and overestimated disconfirmers, McKown and Weinstein (2002) computed hierarchical log-linear models. The results indicated that in the fifth grade, girls were more likely to confirm teachers' underestimations in mathematics and less likely to confirm overestimations than boys. However, such gender differences were found for neither first- and third-grade students in mathematics nor students in any school grade in reading. As McKown and Weinstein (2002) concluded, these results, as well as the lack of significant interactions with continuously measured teacher expectation bias, may be due to the small sample sizes, composed of 121 first graders, 222 third graders, and 218 fifth graders.

In summary, the evidence on the possible moderation of teacher expectancy effects by student gender remains insufficient and inconclusive. Only three studies have examined this issue, and only one study has conducted domain-specific analyses separating the response to positively biased expectations from the response to negatively biased expectations and therefore was able to detect the theoretically assumed response pattern. Moreover, the existing results remain widely inconclusive. Current evidence supports no gender specificity in reading and no gender specificity for most students in mathematics, but there may be gender specificity in mathematics for a special subgroup of students. However, due to the small sample sizes in the study by McKown and Weinstein (2002), it remains possible that boys and girls in general differ slightly in their response to negative and positive bias in reading and mathematics. Thus, further studies expanding the base of evidence on gender-specific teacher expectancy effects are needed. Finally, as none of the studies compared direct gender effects on reading and mathematics achievement gains with and without controlling for biased teacher expectations and their interaction effects, it remains unclear whether and to what extent teacher expectancy effects contribute to domain-specific gender disparities when taking their possible gender specificity into account.

Teacher expectancy effects and the gender gap: a mediation perspective

According to Ludwig (2007), even though direct evidence linking gender gaps to teacher expectancy effects is rare, the compositional evidence for the separate steps among student gender, teacher expectations, and academic achievement supports a contribution to the gender gap on its own. First, empirical evidence reinforces that teacher expectations vary with student gender beyond actual student achievement, and second, biased teacher expectations have been identified as likely self-fulfilling and therefore affecting student achievement gains. The evidence summarised in the previous sections suggests that gender-specific teacher expectation bias may occur, but the results, nevertheless, remain somewhat inconclusive. Therefore, it is worthwhile to conduct direct analyses of the assumed contribution of teacher expectations to gender differences in student achievement.

To date, only two studies have directly investigated the extent to which teacher expectations mediate gender achievement gaps. The study by De Boer et al. (2010) examined the contribution of teacher expectations to gender gaps in an overall indicator of school success. They found no support for the assumption that teacher expectations significantly contributed to gender differences on this measure. However, as gender gaps appear to be domain specific, it is worthwhile to conduct separate analyses on linguistic and mathematical achievement.

Robinson-Cimpian et al. (2014) directly investigated the extent to which prior and current teacher perceptions of students' maths proficiency mediated the gender gap in mathematics. According to their considerations mentioned above, they examined this issue while taking teacher perceptions of students' behaviour and effort into account. Their study was based on data from more than 10,000 first- and third-grade students participating in the Early Childhood Longitudinal Study, Kindergarten (ECLS-K). They conducted various analyses, including traditional mediation models and analyses based on matching approaches. The results consistently indicated mediations by teacher perceptions of students' maths proficiency. However, the mediated proportions differed substantially among the analytic approaches, grade levels, and indicators of teacher perceptions of student proficiency, from under 1% to up to 85%. Therefore, Robinson-Cimpian et al. (2014) were cautious in concluding what exact proportion of gender gaps in mathematics was due to teacher perceptions. Aside from the strength of capturing various analytic approaches, the study had two limitations that are especially relevant in the context of expectancy effects. Both limitations concern the indicator of teacher perceptions of students' maths proficiency. First, teacher perceptions referred to students' current achievement and therefore did not represent teacher expectations of future achievement. Second, and most importantly, for many students in the study, teachers had changed from a "prior" to a "current" measurement point, that is, from kindergarten to first grade and from first to third grade. This arrangement is problematic because the perceptions of one teacher will not shape other teachers' behaviour in class and, therefore, may not initiate a self-fulfilling prophecy. Due to both limitations, the evidence provided by this study was not related to teacher expectancy effects in their essence. Thus, further studies investigating this mediation that are better suited for analysing teacher expectancy effects are needed, as are studies examining this issue in the reading domain.

The current study

Based on these considerations, the current study examined whether the effects of teacher expectations contributed to differences in achievement between boys and girls at the beginning of schooling while considering two possible mechanisms: (a) student gender as a moderator of teacher expectancy effects and (b) teacher expectancy effects as a mediator of gender achievement disparities. The current study took a two-domain perspective and investigated gender differences in reading as well as in mathematics. Very little research has been conducted in this field, especially for reading. Furthermore, the current study considered earlier results supporting teachers' conditional over- and underestimations of boys and girls depending on teacher perceptions of student motivation and work habits, and it included these measures in the models.

Specifically, the following two research questions were analysed:

(1) Does the magnitude of positive and negative teacher expectancy effects differ for boys and girls in reading and mathematics? (moderation perspective)

On the basis of theoretical considerations, we expected gender-specific teacher expectancy effects in reading as well as in mathematics. We expected boys and girls to be more vulnerable in reading and mathematics, respectively, to the effects of negatively biased teacher expectations. Conversely, boys and girls might benefit more strongly in mathematics and reading, respectively, from the effects of positively biased teacher expectations.

(2) Does gender-specific teacher expectation bias mediate the link between student gender and achievement development in reading and mathematics during first grade? (mediation perspective)

In line with the findings by Robinson-Cimpian et al. (2014) in mathematics, we expected gender-specific teacher expectancy effects to partly explain emerging differences in mathematics achievement in first grade between boys and girls. In reading, we assumed a contribution of gender-specific teacher expectations as well. However, this assumption was based on theoretical considerations rather than direct evidence.

Materials and methods

Sample

In the 2013–2014 school year, the research project *Kompetenzerwerb und Lernvoraussetzungen* (KuL; Competence Acquisition and Learning Preconditions; Kristen et al., 2018) was conducted in $N = 39$ primary schools in North Rhine-Westphalia, Germany. The total sample included $N = 1{,}065$ first-grade students from $N = 67$ classrooms.

Classes in which the teachers changed during the school year were excluded from the analyses. Furthermore, we had to exclude one student because of missing gender information. These exclusions left $N = 1{,}025$ students from $N = 64$ classes in $N = 38$ schools. The teachers were predominantly female (94%)[1] and had an average work experience of 12 years ($SD = 8.89$). At the date of school enrolment, the participating students ($n = 532$ boys, $n = 493$ girls) were, on average, 6 years and 6 months old ($SD = 0.33$, $Min = 5$ years and 7 months, $Max = 8$ years and 1 month). On the basis of data from parental interviews, 36% of the children came from families with immigrant status (i.e., at least one parent born abroad). Students' average socioeconomic background, as indicated by the Highest International Socio-Economic Index of Occupational Status (HISEI; Ganzeboom, 2010), amounted to 52.45 ($SD = 19.53$). It is noteworthy that the families of male students participating in this study had a slightly higher HISEI ($M = 53.79$, $SD = 19.50$) than the families of their female classmates ($M = 51.01$, $SD = 19.47$); $t(1023) = 2.28$, $p < .05$, $r = .07$. Furthermore, the boys came from families with immigrant status (32%) less often than the girls (40%); $\chi^2(1) = 6.44$, $p < .05$.

Instruments

Teacher expectations

A few weeks after the students' school enrolment, teachers rated each of the participating students in their class on five items, indicating the expectations they held for each child's achievement in both German language (three items; $\alpha = .94$) and mathematics (two items; $\alpha = .94$). For example, the teachers were asked, "Compared to his/her fellow students, how well do you expect this child to perform at the end of the school year? … in German language/ … in mathematics?". The teacher rated the items on a 5-point scale ranging from 1 = *far below the class average* to 5 = *far above the class average*. Three of the items originated from the BiKS-3-10 study (Artelt, Blossfeld, Faust, Roßbach, & Weinert, 2013), and the other two were developed in the KuL study.

Teacher perceptions of student motivation and work habits

In the teacher questionnaire at the beginning of first grade, teachers also rated their perceptions of each student's motivation and work habits (5-point scale). From the total of eight items ($\alpha = .96$), three items each were related to lessons in German language and mathematics (e.g., "This child truly enjoys learning in German language lessons/maths lessons"), and two items referred to students' general learning behaviour in school (e.g., "This child works very hard in school").

Student achievement and abilities

Students completed various standardised achievement tests in one-to-one sessions supervised by a trained test administrator. At the beginning and at the end of first grade, students completed the computer-based assessment *Fähigkeitsindikatoren Primarschule (FIPS)* (German version of the Performance Indicators in Primary Schools [PIPS]; Bäuerlein et al., 2012). The subscale *phonological awareness* ($\alpha = .82$) consisted of 26 tasks asking the students to repeat pseudo-words, to identify rhymes, and to divide words into syllables. The subscale *reading* ($\alpha = .96$) included 88 tasks covering students' ideas of reading as well as first reading tasks. The subscale *mathematics* ($\alpha = .92$) comprised 53 tasks. First, students had to solve easy addition and subtraction tasks on the basis of pictures and read-aloud numbers. Subsequently, students were asked to perform calculations on a more abstract level based on dots and in formal arithmetic problems. In all subscales, the tasks were sorted by degree of difficulty and were given to a student until a predefined number of mistakes had been reached. Depending on students' achievement level, the instrument captured predominantly precursor skills or first skills in reading and mathematics. In each domain, a sum score was generated, with one point awarded for each solved task.

At the beginning of the school year, students also completed the matrix test of the *Grundintelligenztest Skala 1 (CFT1)* (German version of the Culture Fair Intelligence Test; Weiß & Osterland, 1997; $\alpha = .78$), which captured deductive reasoning through 12 items, and the subscale *working memory*, which was implemented in the FIPS assessment (Bäuerlein et al., 2012; 10 items; $\alpha = .76$). The two scales depicted students' general cognitive abilities.

Student motivation and self-concept

All participating children were interviewed in one-to-one sessions about their joy of learning and effort (middle of the school year) as well as their academic self-concept (beginning of first grade). Students' joy of learning (13 items; $\alpha = .78$) and effort (13 items; $\alpha = .70$) were measured with an adapted form of the *Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrungen von Grundschulkindern erster und zweiter Klassen* (FEESS1-2) (Questionnaire for emotional and social school experiences of first- and second-grade students; Rauer & Schuck, 2004). For each statement regarding joy of learning (e.g., "I like to learn at school") or effort (e.g., "I also try to solve very difficult tasks"), the children indicated whether it applied to them on a scale with 0 = *not true*, 1 = *partly true*, 2 = *completely true*. Students rated their academic self-concept on questions developed by Poloczek, Greb, and Lipowsky (2009) in reading (6 items; $\alpha = .62$) and mathematics (5 items; $\alpha = .75$) using a 3-point scale (e.g., "How do you perform in reading/ in calculating?" 1 = *not well*, 2 = *well*, 3 = *very well*).

Student gender and background characteristics

Students' gender was captured as a dummy variable in the analyses (0 = male, 1 = female). We also considered three aspects of students' family background: socioeconomic status, education level, and immigrant status. Whereas the HISEI among the parents

(Ganzeboom, 2010) was used to account for parental socioeconomic status, parental education was captured by a dummy-coded variable differentiating between families with at least one parent with *Abitur* (higher education entrance qualification in Germany), coded as 1, and without *Abitur*, coded as 0. We also took into account whether the students came from immigrant families, which was also captured with a dummy variable (0 = non-immigrant family, 1 = immigrant family). A student was assigned to the immigrant family group if at least one parent was born abroad. Furthermore, students' age in months at the time of school enrolment served as a control variable.

Data preparation

General information and imputation

All analyses were conducted using Stata 14.2 (StataCorp LLC, 1985–2015). Variables with missing information were imputed under the missing-at-random assumption using the fully conditional specification (Van Buuren, Brand, Groothuis-Oudshoorn, & Rubin, 2006). The imputation models included not only the analytic variables but also further information collected within teacher and parent interviews. We conducted all analyses individually for the 50 imputed datasets and pooled the parameters according to Rubin's rules (Rubin, 1987). As classroom effects were not the focus of our study, we did not conduct hierarchical linear models; rather, we used robust standard errors in all regression analyses to take the clustered data structure (students within classrooms) into account. Descriptive statistics and preliminary analyses were based on the first imputed dataset.

Separating teacher expectation bias

As only biased teacher expectations may result in self-fulfilling prophecies (e.g., Jussim et al., 2009), we first sought to separate the proportion of bias in teacher expectations. To identify such bias, we followed the residual approach proposed by Madon, Jussim, and Eccles (1997). We conducted multiple regression analyses predicting teacher expectations in the German language and mathematics from the following student variables: phonological awareness, reading and mathematics achievement; cognitive abilities; motivation; and academic self-concept. The residuals resulting from these regressions reflect the variance that is unexplained by the student characteristics mentioned and can thus be understood as teacher expectation bias. Residual scores close to zero indicated unbiased expectations, positive values indicated positively biased expectations, and negative values indicated negatively biased expectations.

Categorisation of teacher expectation bias and student achievement

To replicate the models by McKown and Weinstein (2002), we further divided the linear variable of teacher expectation bias into categories of bias distinguishing between positive bias (residual score more than 0.3 standard deviations above the mean), negative bias (residual score more than 0.3 standard deviations below the mean), and no bias (residual score within 0.3 standard deviations of the mean). Following the considerations by McKown and Weinstein (2002), we used the cut-off criterion of 0.3 standard deviations, which indicates a medium-sized effect in the typology of effects sizes by Cohen (1988). In the next step, we categorised the students' achievement responses. Therefore, we predicted student end-of-year achievement based on the students' prior achievement, cognitive abilities, motivation, and self-concept. The resulting residual scores depicted the extent to which end-of-year achievement deviated from what one would predict based on students' prior skills and motivation. In line with the procedure of categorising teacher expectation bias, we divided these achievement residuals into three categories: improved (residual score more than 0.3 standard deviations above the mean), declined (residual score more than 0.3 standard deviations below the mean), and remained stable (residual score within 0.3 standard deviations of the mean). Subsequently, we combined the two categorisations, identifying students whose achievement developed in the direction of teachers' biased expectations. These students were labelled confirmers (positive bias and improved achievement or negative bias and declined achievement). In the cases in which student achievement developed in the opposite direction of teacher expectations bias (positive bias and declined achievement or negative bias and improved achievement), the students were assigned to the group of disconfirmers. Students who were targets of no bias or whose achievement remained stable were not further categorised and were excluded from the subsequent analyses.

Testing the hypotheses

Gender-specific teacher expectancy effects: moderation assumption

We first examined the question regarding gender-specific responses to teacher expectations (Research Question 1). Therefore, we conducted separate analyses for reading and for mathematics. We began by modelling three-way interactions among student gender, teacher expectation bias, and squared teacher expectation bias in predicting students' end-of-year achievement, controlling for prior student achievement, cognitive abilities, age, and family characteristics. The squared term of teacher expectation bias was included to model non-linear relationships.

Second, seeking to replicate models by McKown and Weinstein (2002), we investigated whether male and female students differed in their likelihood of confirming or disconfirming teachers' biased expectations as a function of whether the expectations were negatively or positively biased. Therefore, we used the categorisations of teacher expectation bias (negative bias and positive bias) and student responses (confirming or disconfirming) and conducted logistic regression analyses predicting students' likelihood of confirming or disconfirming teacher expectation bias as a function of student gender, and positive versus negative teacher expectation bias, controlling for students' age and family background.

Contribution to the gender gap: mediation assumption

After investigating gender specificity in teacher expectancy effects, we evaluated the extent to which the effects of teacher expectations contributed to gender gaps in students' early achievement (Research Question 2). Therefore, we first calculated linear regression models predicting students' end-of-year achievement based on student gender, age, and family characteristics (socioeconomic status, educational level, and immigrant status). In a second model, we added students' beginning-of-year achievement and measures of their general cognitive abilities. This second model captured a calculation of gender gaps in students' end-of-year achievement that existed beyond differences in students' prior achievement, general cognitive abilities, age, and family characteristics. This model and a third model that additionally controlled for teacher perceptions of student motivation and work habits served as references for the final model, which also included teacher expectation bias. A comparison of the magnitude of the gender effect from these models would provide evidence of potential mediation by teachers' biased expectations. In the case that prior analyses provided support for gender-specific teacher expectancy effects, we included relevant interaction terms in the final model.

Results

Preliminary analyses

Table 1 shows descriptive statistics for all variables separately for boys and girls. As the mean differences in the residual scores of teacher expectations indicated, teachers expected higher reading achievement for girls ($M = 0.13$, $SD = 0.76$) than for boys ($M = -0.11$, $SD = 0.77$), accounting for students' actual achievement but not for differences in teacher perceptions of student motivation and work habits; $t(1023) = -5.19$, $p < .001$, $r = .16$; that is, girls were more likely to be overestimated by their teachers in reading achievement, whereas boys became the target of negatively biased expectations more frequently, when teacher perceptions of student motivation and work habits were not taken into account. In mathematics, no gender differences in teacher expectation bias (girls: $M = -0.02$, $SD = 0.69$, boys: $M = 0.02$, $SD = 0.67$) occurred; $t(1001) = 1.11$, $p = .27$, $r = .04$. However, opposing findings emerged when teacher perceptions of student motivation and work habits were also accounted for. With regard to similarly performing students whom the teacher perceived as being equally motivated and similarly behaved in school, boys became targets of higher expectations in mathematics than girls; $t(1001) = -4.43$, $p < .001$, $r = .14$. In reading, the gender difference decreased when teacher perceptions of student motivation and work habits were accounted for; $t(1023) = 2.17$, $p < .05$, $r = .07$. For more detailed information on group differences in teacher expectations and conditional over- and underestimations within these data, see Lorenz et al. (2016) and Gentrup, Rjosk, et al. (2018).

On the basis of the categories of teacher expectation bias, teachers held unbiased expectations towards 25% to 28% of their male and female students. In both domains, positive and negative bias varied between 30% and 42% for boys and girls (see Table 1). As revealed by *t* tests on the underlying continuous variables of teacher expectation bias, the scores were significantly different from zero and therefore indicated negative or positive bias (see Appendix 1).

Furthermore, the descriptive results indicated some differences in achievement between boys and girls. Whereas male and female students showed similar phonological awareness and reading skills at the beginning and at the end of first grade (all *t* tests were non-significant), boys ($M = 28.81$, $SD = 8.41$) were already outperforming girls ($M = 25.06$, $SD = 7.19$) in mathematics at the beginning of first grade; $t(1023) = 7.64$, $p < .001$, $r = .23$. This male advantage persisted with a similar magnitude at the end of first grade; $t(1023) = 8.54$, $p < .001$, $r = .26$.

Gender-specific teacher expectancy effects: moderation perspective

To investigate gender specificity in teacher expectancy effects, we conducted two different types of analyses. First, we investigated the interactions among students'

Table 1. Descriptive statistics of all variables for the total sample as well as for female and male students.

	Total Sample[a]				Female[b]		Male[c]	
Variable	*M*	*SD*	*Min*	*Max*	*M*	*SD*	*M*	*SD*
Student Beginning-of-Year Achievement								
Phonological Awareness	13.22	4.66	1	26	13.31	4.56	13.14	4.75
Reading	23.29	26.39	0	156	23.22	24.25	23.36	28.25
Mathematics	27.01	8.06	0	53	25.06	7.19	28.81*	8.41
Student End-of-Year Achievement								
Phonological Awareness	20.78	4.65	3	26	20.99	4.34	20.59	4.91
Reading	107.36	37.79	3	158	108.99	36.61	105.85	38.84
Mathematics	38.60	7.06	17	53	36.71	6.23	40.35*	7.33
Student Cognitive Abilities								
Deductive Reasoning	6.59	2.92	0	12	6.82	2.84	6.38*	2.98
Working Memory	3.11	1.70	0	8	3.02	1.60	3.20	1.78
Student Motivation and Self-Concept								
Joy of Learning	1.53	0.38	0	2	1.58	0.35	1.49*	0.40
Effort	1.69	0.28	0.31	2	1.69	0.27	1.69	0.30
Academic Self-Concept Reading	1.84	0.54	1	3	1.86	0.55	1.81	0.53
Academic Self-Concept Mathematics	2.25	0.44	1	3	2.22	0.46	2.28	0.42
Teacher Expectation								
German Language	3.31	0.98	1	5	3.41	0.97	3.23*	0.97
Mathematics	3.36	0.88	1	5	3.27	0.87	3.45*	0.88
Teacher Expectation Bias								
German Language	0.00	0.77	−2.75	2.06	0.13	0.76	−0.11*	0.77
Mathematics	0.00	0.68	−2.43	2.41	−0.02	0.69	0.02	0.67
Teacher Perception								
Motivation and Work Habits	3.66	0.87	1	5	3.83	0.86	3.50*	0.86
Student Background								
Age	6.46	0.33	5.58	8.08	6.44	0.32	6.47	0.33
HISEI	52.45	19.53	14.21	88.96	51.01	19.47	53.79*	19.50
			N	$\%_{valid}$	*n*	$\%_{valid}$	*n*	$\%_{valid}$
Parental Education								
No Abitur			510	49.76	261	52.94	249	46.80
Abitur or higher			515	50.24	232	47.06	283	53.20
Immigrant Status								
No			660	64.39	298	60.45	362	68.05
Yes			365	35.61	195	39.55	170	31.95
Categories of Teacher Expectation Bias								
German Language								
Negative Bias			368	35.90	150	30.43	218	40.98
No Bias			267	26.05	136	27.59	131	24.62
Positive Bias			390	38.05	207	41.99	183	34.40
Mathematics								
Negative Bias			369	36.79	193	39.71	176	34.04
No Bias			265	26.42	123	25.31	142	27.47
Positive Bias			369	36.97	170	34.98	199	38.49
Categories of Student Response								
Reading								
Disconfirming Achievement			218	21.27	110	22.31	108	20.30
Confirming Achievement			372	36.29	168	34.08	204	38.35
Uncategorised			435	42.44	215	43.61	220	41.35
Mathematics								
Disconfirming Achievement			212	20.68	113	22.92	99	18.61
Confirming Achievement			304	29.66	150	30.43	154	28.95
Uncategorised			509	49.66	230	46.65	279	52.44

Note: All descriptive statistics are based on the first imputed dataset. HISEI = Highest International Socio-Economic Index of Occupational Status. *The mean difference between female and male students is statistically significant at the 5% significance level (independent *t* tests).

[a]*N* = 1,025. [b]*n* = 493. [c]*n* = 532.

gender, teacher expectation bias, and squared teacher expectation bias when predicting students' end-of-year achievement. Second, we examined gender differences in confirming and disconfirming teacher expectations.

Reading

An investigation of continuously measured teacher expectation bias revealed no significant interaction terms between student gender and either teacher expectation bias or squared teacher expectation bias (see Appendix 2). In other words, these analyses did not provide support for gender differences in teacher expectancy effects for boys and girls in reading.

Equally, when examining boys' and girls' likelihood of confirming or disconfirming teacher expectations as a function of positive and negative bias, no gender differences emerged ($b = 0.49$, $p = .15$; see Appendix 3). According to these analyses, boys and girls were equally likely to confirm (or disconfirm) teachers' biased expectations in reading ($b = -0.42$, $p = .14$). Furthermore, students' sensitivity towards teacher expectations appeared to be independent of whether teacher expectations were positively or negatively biased ($b = 0.05$, $p = .88$).

Mathematics

The results investigating continuously measured teacher expectation bias in mathematics were similar to the results for reading. No significant interaction between student gender and teacher expectation bias or between student gender and squared teacher expectation bias occurred (see Appendix 4).

However, the logistic models investigating boys' and girls' differential responses to teachers' positively or negatively biased expectations showed significant results. When predicting students' likelihood of confirming teachers' biased expectations in mathematics, a significant interaction effect between student gender and positive versus negative bias ($b = -1.44$, $p < .001$; see Appendix 5), as well as a significant main effect of gender ($b = 0.61$, $p < .01$), occurred while controlling for students' age and family background. The main effect of positive bias was not significant ($b = 0.52$, $p = .12$). Figure 1 illustrates the differing response patterns for boys and girls in mathematics. As post-hoc analyses comparing student responses within and between genders revealed, girls were more likely to confirm negatively biased teacher expectations than were boys (effect of comparison between girls being targets of negative bias and boys being targets of negative bias: $b = 0.61$, $p < .01$). Conversely, boys were more likely to confirm positively biased teacher expectations than were girls ($b = 0.83$, $p < .01$). Furthermore, girls were generally less likely to confirm positively biased than negatively biased teacher expectations ($b = -0.92$, $p < .05$). For boys, the difference between their response to positively and negatively biased teacher expectations was non-significant ($b = 0.52$, $p = .12$).

In summary, the analyses that included the whole sample revealed no gender specificity in teacher expectancy effects for reading or mathematics. As a result, we did not include any interaction terms in further analyses investigating the contribution of teacher expectancy effects to gender achievement gaps for the overall sample.

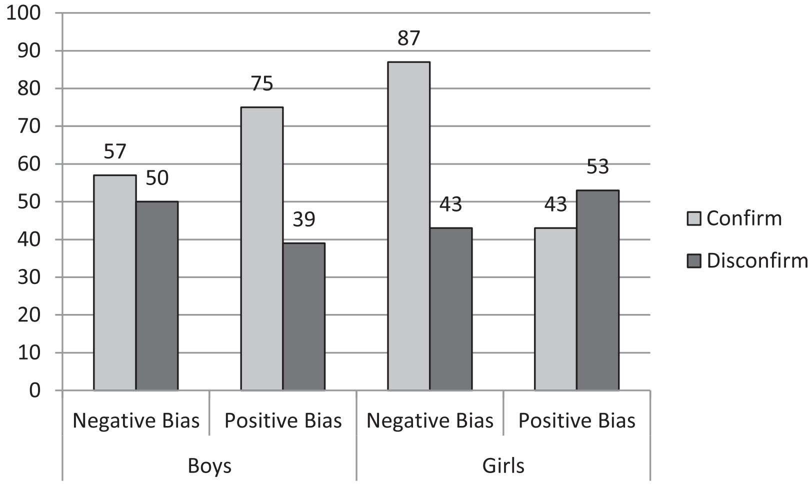

Figure 1. Absolute frequencies of negative and positive teacher expectation bias by response pattern in mathematics achievement by gender based on the first imputed dataset.

Gender gap and teacher expectations: mediation perspective

To investigate the extent to which teacher expectancy effects contributed to gender gaps in reading and mathematics (Research Question 2), we calculated four stepwise models that included further predictors of student end-of-year achievement.

Reading

As Model 1 in Table 2 indicates, female students showed somewhat higher reading skills at the end of first grade than their male classmates ($\beta = .14$, $p < .05$) when taking differences in their age and family background into account. This female advantage decreased ($\beta = .08$, $p = .11$) when students' beginning-of-year phonological awareness, reading skills, and general cognitive abilities were also taken into account. Thus, even though female students overall outperformed male students in reading at the end of first grade, they did not show greater achievement gains during the year. The subsequent addition of teacher perceptions of student motivation and work habits and teacher expectation bias to the models further reduced the regression coefficient of student gender to values close to zero. However, this reduction was not statistically significant (both $\Delta\beta$ non-significant); that is, teacher expectation bias did not substantially mediate the direct link between student gender and end-of-year reading achievement. Further, it is remarkable that the prediction of students' reading gains by teacher expectation bias decreased (from $\beta = .16$, $p < .001$ [model not displayed; for more details, see Gentrup, Lorenz, et al., 2018] to $\beta = .08$, $p = .06$) if teacher perceptions of student motivation and work habits were included in the model. However, teacher perceptions of student motivation and work habits remained a significant predictor of student reading gains ($\beta = .17$, $p < .001$).

Mathematics

In mathematics, a significant gender difference in students' end-of-year achievement emerged (see Table 3). On average, taking students' age and family background into

Table 2. Teacher expectancy effects and the gender gap in reading.

	Model 1		Model 2		Model 3			Model 4		
Variable	*b* (*SE*)	*β*	*b* (*SE*)	*β*	*b* (*SE*)	*β*	*Δβ*	*b* (*SE*)	*β*	*Δβ*
Student Gender										
Female	5.14 (2.08)	.14*	3.16 (1.92)	.08	0.03 (1.99)	.00	.08	−0.49 (2.03)	−.01	.09
Student Background										
Student Age	5.60 (3.79)	.05	−0.25 (3.75)	−.00	−1.00 (3.71)	−.01		−0.82 (3.66)	−.01	
HISEI	0.44 (0.08)	.23***	0.24 (0.07)	.12**	0.20 (0.07)	.10*		0.18 (0.07)	.09*	
Abitur	4.48 (3.02)	.12	2.00 (2.87)	.05	1.20 (2.88)	.03		0.95 (2.88)	.03	
Immigrant Status	−2.32 (2.96)	−.06	−1.84 (2.64)	−.05	−1.98 (2.62)	−.05		−1.41 (2.66)	−.04	
Student Beginning-of-Year Achievement										
Phonological Awareness			2.08 (0.28)	.26***	1.74 (0.27)	.22***		1.86 (0.28)	.23***	
Reading			0.06 (0.04)	.04	0.03 (0.04)	.02		0.04 (0.04)	.03	
Student Cognitive Abilities										
Deductive Reasoning			2.21 (0.43)	.17***	1.87 (0.41)	.14***		1.97 (0.41)	.15***	
Working Memory			1.59 (0.59)	.07**	0.93 (0.56)	.04		1.03 (0.56)	.05	
Teacher Perceptions										
Student Motivation and Work Habits					9.22 (1.49)	.22***		7.23 (1.66)	.17***	
Teacher Expectation Bias										
German language								4.05 (2.11)	.08	
Intercept	44.47 (25.66)		46.06 (25.13)		31.04 (24.96)			35.32 (25.20)		
R^2	7.54%		21.97%		25.54%			26.02%		
Cohen's f^2			0.18		0.05			0.01		

Note: HISEI = Highest International Socio-Economic Index of Occupational Status. *Δβ* = difference in beta-coefficient for the female predictor compared to Model 2. Cohen's f^2 = effect size of change in R^2 compared to the prior model. *N* = 1,025. **p* < .05. ***p* < .01. ****p* < .001.

account, boys showed approximately 0.47 standard deviations greater mathematics skills at the end of first grade than their female classmates ($\beta = -.47$, $p < .001$). After additionally controlling for students' prior mathematics achievement and cognitive abilities, this advantage for boys decreased but remained statistically significant ($\beta = -.23$, $p < .001$). Therefore, on average, boys gained 0.23 standard deviations more mathematics skills in first grade than their female classmates who had similar beginning-of-year achievement, cognitive abilities, age, and family characteristics. Furthermore, teacher expectation bias significantly predicted students' end-of-year achievement in mathematics ($\beta = .07$, $p < .001$). Nevertheless, the advantage of boys did not substantially change when teacher expectation bias was added to the model (*Δβ* non-significant); that is, the stronger achievement gains of boys in mathematics seem to occur independently of teacher expectations and their effects on end-of-year mathematics achievement.

Table 3. Teacher expectancy effects and the gender gap in mathematics.

	Model 1		Model 2		Model 3			Model 4		
Variable	*b* (*SE*)	*β*	*b* (*SE*)	*β*	*b* (*SE*)	*β*	*Δβ*	*b* (*SE*)	*β*	*Δβ*
Student Gender										
Female	−3.29 (0.44)	−.47***	−1.63 (0.33)	−.23***	−1.89 (0.36)	−.27***	.04	−1.67 (0.36)	−.24***	.01
Student Background										
Student Age	1.08 (0.63)	.05	−0.50 (0.47)	−.02	−0.54 (0.47)	−.03		−0.56 (0.47)	−.03	
HISEI	0.07 (0.01)	.20***	0.02 (0.01)	.05	0.02 (0.01)	.05		0.02 (0.01)	.05	
Abitur	0.95 (0.50)	.14	0.54 (0.35)	.08	0.48 (0.35)	.07		0.44 (0.34)	.06	
Immigrant Status	−1.15 (0.51)	−.17*	−0.10 (0.36)	−.01	−0.14 (0.36)	−.02		−0.18 (0.36)	−.03	
Student Beginning-of-Year Achievement										
Mathematics			0.55 (0.03)	.63***	0.53 (0.03)	.61***		0.55 (0.03)	.63***	
Student Cognitive Abilities										
Deductive Reasoning			0.27 (0.06)	.11***	0.26 (0.06)	.11***		0.27 (0.06)	.11***	
Working Memory			0.09 (0.09)	.02	0.06 (0.09)	.02		0.08 (0.09)	.02	
Teacher Perceptions										
Student Motivation and Work Habits					0.52 (0.25)	.06*		0.15 (0.28)	.02	
Teacher Expectation Bias										
Mathematics								0.76 (0.28)	.07**	
Intercept	29.35*** (4.14)		24.42*** (3.23)		23.78*** (3.17)			24.64*** (3.16)		
R^2	13.81%		56.79%		57.20%			57.48%		
Cohen's f^2			0.99		0.01			0.01		

Note: HISEI = Highest International Socio-Economic Index of Occupational Status. *Δβ* = difference in beta-coefficient for the female predictor compared to Model 2. Cohen's f^2 = effect size of change in R^2 compared to the prior model. *N* = 1,003. *$p < .05$. **$p < .01$. ***$p < .001$.

Discussion

The current study investigated the role of teacher expectations in early gender disparities by analysing two complementary research questions. First, we examined whether boys and girls differed in their vulnerability to teacher expectancy effects. Second, we analysed whether gender-specific effects of teacher expectations contributed to the emerging gender gap in reading and mathematics in the first year of schooling.

Gender-specific teacher expectancy effects: moderation perspective

Analyses on the first research question revealed no support for gender-specific teacher expectancy effects for the overall student sample in reading or mathematics. In other words, considering all participating students, negatively and positively biased teacher expectations affected boys' and girls' learning to a similar extent. Further analyses on this research question focused on a subgroup of students. These students were targets of relatively strong bias in teacher expectations, and their achievement development differed from what was expected based on their prior achievement, cognitive abilities, motivation, and self-concept. In this specific subgroup, girls in mathematics were, as hypothesised, more affected by negatively biased expectations than by positively biased expectations and were more vulnerable to negative teacher expectation bias than boys. For boys, the likelihood of confirming teacher expectation bias did not vary significantly with the value of bias, but it showed a higher tendency for positive bias. In other words, and in line with our assumptions, in this specific subgroup of students, girls were shown to be more vulnerable to negative bias and to be less supported by positive bias in mathematics, whereas boys tended to be more likely to benefit from positive bias than to suffer from negative bias. In reading, no such differences in boys' and girls' vulnerability to expectancy effects was apparent.

Overall, our result patterns were in accordance with the findings of McKown and Weinstein (2002). First, both studies identified no gender difference in teacher expectancy effects in the overall student sample. Second, in both studies, gender-specific responses to teacher expectation bias for students who were targets of relatively strong bias occurred only in mathematics. However, with regard to students' grade level, our results, which were based on first-grade students, were not in line with those of McKown and Weinstein (2002), who found gender-specific response patterns in fifth-grade students but not in first- or third-grade students. This finding may have arisen due to the small number of first-grade students in their sample compared to our sample.

In comparing our results and those of McKown and Weinstein (2002) with the findings of Jussim et al. (1996), it is essential to distinguish between the indicators of teacher perceptions considered as well as between the two mathematics-related outcome variables. All three studies consistently found no support for overall gender differences in teacher expectancy effects in mathematics when examining mathematics test scores as a dependent variable. However, further findings by Jussim et al. (1996) indicated that this result did not emerge for maths grades as an outcome. Here, stronger effects of teacher perceptions of students' maths talent were observed for girls than for boys. As final maths grades are, strictly speaking, teacher judgements themselves, this result can be interpreted as a stronger overlap between two forms of teacher judgement for girls than for boys.

Summarising the results, we conclude that in reading, teacher expectancy effects were independent of student gender.[2] This also seems true for most students in mathematics. However, in mathematics, teacher expectations may reinforce gender-stereotypic achievement development for the student group with the strong bias in teacher expectations mentioned above.

Further studies should focus on this special group, which is not small in number (44% of the participating students in our study), and investigate the characteristics of students belonging to this group. Do they possibly perform at the extremes of the mathematics achievement distribution? Are students from different social backgrounds equally represented? Additionally, it seems worthwhile to conduct further analyses on the gender-specific mechanisms linking negatively biased expectations to disadvantages in girls' learning and positively biased expectations to advantages in boys' development.

Gender gap and teacher expectations: mediation perspective

The results regarding the second research question, which investigated the extent to which teacher expectancy effects contributed to gender achievement gaps, contradicted our assumptions. Neither in reading nor in mathematics did teacher expectancy effects substantially contribute to gender gaps in first-grade students. In conclusion, according to our findings, the overall impact of the effects of teacher expectations on emerging gender gaps was negligible.

Therefore, our findings contradicted the results presented by Robinson-Cimpian et al. (2014). As the data in our study covered teacher expectations measured prior to student achievement and from teachers actually working with the students, we directly investigated teacher expectancy effects in terms of a self-fulfilling prophecy. Thus, the underlying data were better suited to examine the contribution of teacher expectancy effects to gender gaps. Further, aside from issues of domain specificity, our results were in accordance with the findings of De Boer et al. (2010).

However, as our study investigated effects only within the short time period of almost one school year (on average, 8 months passed from the first to last measurement points), one may assume that a meaningful contribution of teacher expectations to gender gaps in mathematics may still be possible in the long term. This assumption would be compatible with the results by Robinson-Cimpian et al. (2014), whose models covering a 1-year period (kindergarten to first grade) showed somewhat smaller effect sizes than models covering 2 years (first grade to third grade). However, De Boer et al. (2010) investigated mediation in overall performance over a 5-year period and did not find support for a substantial contribution to the gender gap. In conclusion, it would be possible, though somewhat unlikely, that substantial mediation effects in mathematics might become salient only over longer time periods. However, given the results on girls' greater vulnerability to negatively biased teacher expectations in the subgroup of students experiencing relatively strong bias in teacher expectations, stronger contributions to the gender gap in mathematics may be found in this subgroup. Further research on this issue would be worthwhile.

For reading, our study revealed no gender gaps in students' achievement development during first grade. Accordingly, teacher expectancy effects did not significantly contribute to a gender gap in reading. Thus, additional studies in later school years, when gender-related reading disparities become stronger (OECD, 2016), should investigate whether teacher expectancy effects play a more important role later in this process.

As the current study and the research by McKown and Weinstein (2002) observed no gender-specific responses to teacher expectations in reading, it seems reasonable to assume that teacher expectancy effects in the early school years are less important for gender gaps in reading than they may be in mathematics. It is possible that gender stereotypes stigmatising girls in mathematics are more common than stereotypes disadvantaging boys in reading. To our knowledge, there is no research directly comparing the prevalence of the girls' maths stereotype to the boys' reading stereotype. In the context of stereotype lift, Walton and Cohen (2003) argued that men (and Whites) are seen as ordinary in most domains of society and that they are thus less likely to be targets of negative or positive stereotypes. Additionally, the overall small number of published studies on stereotypes in the reading domain might indicate that the boys' reading stereotype is less prevalent. However, further investigation is needed to provide more information on this issue.

Limitations and future research

The present study has some limitations that should be addressed in future research. First, this study did not examine the role of teachers' gender. As almost all participating teachers were female, the results are transferable only to classrooms taught by female teachers. Thus, the role of teachers' gender in the transmission of teacher expectations remains unclear. It might be argued that a gender match between student and teacher may reduce the threat originating from negatively biased teacher expectations, as the lower estimations stem from an in-group judge (cf. McKown & Weinstein, 2002). In contrast, one may argue that a gender match strengthens the negative effects because students identify themselves more strongly with same-sex teachers, whom they may see as role models. Although our study did not focus on this issue, the results speak to the latter assumption. In our study, female students who were taught by female maths teachers were more adversely affected by negatively biased expectations than were boys. In contrast, no specific disadvantages for boys who were taught by female teachers in reading occurred. Further studies should examine this issue in more detail.

Second, in this study, it was not possible to entirely conclude whether the observed relationships between teacher expectation bias and end-of-year achievement were due to causal effects. The longitudinal study design, which was characterised by minimal teacher–student contact before measuring teacher expectations and students' prior achievement as well as by various measurements of student characteristics, such as general cognitive abilities and different indicators of students' family background, reduced the number of possible influences that may have provided alternative explanations. However, influences of other variables not considered in this study are possible.

Further, the interplay between teacher expectation bias and teacher perception of student motivation and work habits in predicting student achievement is subject to discussion and raises some issues for future research. Both teacher judgements substantially predicted student end-of-year achievement when prior achievement, general cognitive ability, and background characteristics were controlled for. However, when both teacher judgements were simultaneously included in the models, in reading, only teacher perception of student motivation and work habits remained a significant predictor. This result supports the assumption that teachers' perceptions of their students' motivation and work habits may have an impact on student achievement, at least in reading, beyond their achievement expectations. However, the substantial overlap between the two measures, indicating that students who were assumed to work hard were also more likely to become targets of overestimations, makes it difficult to determine cause–effect relationships. Further, this phenomenon occurred only in reading, which may be because in our study, all reading teachers were class teachers. Those teachers are strongly involved in instructing first-grade students on school and class rules. In contrast, some of the maths teachers were responsible only for maths lessons. However, this did not apply to all maths teachers. Future research should try to disentangle the effects of the two teacher judgements – especially in the early years of schooling, when compliant school behaviour is a major educational goal – while investigating student motivation and work habits as well as student achievement in reading and mathematics as outcomes.

In conclusion, the current study indicated that, overall, gender-specific teacher expectations did not contribute to the gender gap, especially for reading outcomes, during the first year of schooling. However, the finding that some students were more vulnerable to teachers' over- or underestimations of mathematics achievement calls for future studies in this domain.

Notes

1. We also conducted all analyses on the sample of female teachers and their students; in other words, we excluded male teachers and their students from the analyses. All results were equal to the reported results for the overall sample.
2. Teacher expectation bias significantly predicted end-of-year reading achievement only when teacher perceptions of student motivation and work habits were not taken into account. This finding was due to a substantial overlap between the two measures, indicating that students who were assumed to work hard were more likely to become targets of overestimations. However, as substantial effects of teacher expectation bias occurred when entered separately into the models, the results indicated teacher expectancy effects in reading as well.

Acknowledgement

The authors acknowledge financial support by the Federal Ministry of Education and Research of Germany [grant number 01JC1117]. The responsibility for the content of this publication lies with the authors. We acknowledge support by the Open Access Publication Fund of Humboldt-Universität zu Berlin.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Federal Ministry of Education and Research of Germany [grant number 01JC1117].

Notes on contributors

Sarah Gentrup is a doctoral candidate at the Institute for Educational Science at Humboldt-Universität zu Berlin. Her research interests are teacher expectations, teacher–student interactions, the achievement development of young students, and the formation of gender-related educational disparities.

Camilla Rjosk is a postdoctoral research scientist at the German Institute for Educational Quality Improvement at Humboldt-Universität zu Berlin. Her research interests encompass teacher expectations, their beliefs and instructional behaviour, classroom composition related to diversity in student backgrounds, and the achievement development of disadvantaged student groups.

References

Anders, Y., McElvany, N., & Baumert, J. (2010). Die Einschätzung lernrelevanter Schülermerkmale zum Zeitpunkt des Übergangs von der Grundschule auf die weiterführende Schule: Wie differenziert urteilen Lehrkräfte? [Perception of learning relevant student characteristics at school transition. How differentiated are teacher judgements?]. In K. Maaz, J. Baumert, C. Gresch, & N. McElvany (Eds.), *Der Übergang von der Grundschule in die weiterführende Schule: Leistungsgerechtigkeit und* *regionale, soziale und ethnisch-kulturelle Disparitäten* (pp. 313–330). Bonn: Bundesministerium für Bildung und Forschung, Referat Bildungsforschung.

Artelt, C., Blossfeld, H.-P., Faust, G., Roßbach, H.-G., & Weinert, S. (2013). *Bildungsprozesse, Kompetenzentwicklung und Selektionsentscheidungen im Vorschul- und Schulalter (BiKS-3-10)* [Educational processes, competence development and selection decisions in preschool and school age (BiKS-3-10)]. (Version 2) [Data set]. Berlin: IQB – Institut zur Qualitätsentwicklung im Bildungswesen. doi:10.5159/IQB_BIKS_3_10_v4

Bäuerlein, K., Beinicke, A., Berger, N., Faust, G., Jost, M., & Schneider, W. (2012). *FIPS Fähigkeitsindikatoren Primarschule: Ein computerbasiertes Diagnoseinstrument zur Erfassung der Lernausgangslage und der Lernentwicklung von Schulanfängern. Manual* [Performance indicators in primary schools: A computer-based diagnostic tool for detecting the initial learning capabilities and the learning development of children starting school]. Göttingen: Hogrefe.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

De Boer, H., Bosker, R. J., & Van der Werf, M. P. C. (2010). Sustainability of teacher expectation bias effects on long-term student performance. *Journal of Educational Psychology*, *102*, 168–179. doi:10.1037/a0017289

Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology*, *53*, 25–44. doi:10.1016/j.jsp.2014.10.002

Friedrich, A., Flunger, B., Nagengast, B., Jonkmann, K., & Trautwein, U. (2015). Pygmalion effects in the classroom: Teacher expectancy effects on students' math achievement. *Contemporary Educational Psychology*, *41*, 1–12. doi:10.1016/j.cedpsych.2014.10.006

Fryer, R. G., Jr., & Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, *2*(2), 210–240. doi:10.1257/app.2.2.210

Ganzeboom, H. B. G. (2010, May). *A new international socio-economic index (ISEI) of occupational status for the international standard classification of occupation 2008 (ISCO-08) constructed with data from the ISSP 2002–2007*. Paper presented at the Annual Conference of the International Social Survey Programme, Lisbon, Portugal.

Gentrup, S., Lorenz, G., Kristen, C., & Kogan, I. (2018). *Self-fulfilling prophecies in the classroom: Teacher expectations, teacher feedback and student achievement*. Manuscript submitted for publication.

Gentrup, S., Rjosk, C., Stanat, P., & Lorenz, G. (2018). Einschätzungen der schulischen Motivation und des Arbeitsverhaltens durch Grundschullehrkräfte und deren Bedeutung für Verzerrungen in Leistungserwartungen [Teachers' perceptions of students' motivation and learning behaviour and their role in biased teacher achievement expectations]. *Zeitschrift für Erziehungswissenschaft*, *21*, 867–891. doi:10.1007/s11618-018-0806-2

Glock, S. (2016). Stop talking out of turn: The influence of students' gender and ethnicity on preservice teachers' intervention strategies for student misbehavior. *Teaching and Teacher Education*, *56*, 106–114. doi:10.1016/j.tate.2016.02.012

Glock, S., & Kleen, H. (2017). Gender and student misbehavior: Evidence from implicit and explicit measures. *Teaching and Teacher Education*, *67*, 93–103. doi:10.1016/j.tate.2017.05.015

Heyder, A., & Kessels, U. (2013). Is school feminine? Implicit gender stereotyping of school as a predictor of academic achievement. *Sex Roles*, *69*, 605–617. doi:10.1007/s11199-013-0309-9

Hinnant, J. B., O'Brien, M., & Ghazarian, S. R. (2009). The longitudinal relations of teacher expectations to achievement in the early school years. *Journal of Educational Psychology*, *101*, 662–670. doi:10.1037/a0014306

Holder, K., & Kessels, U. (2017). Gender and ethnic stereotypes in student teachers' judgments: A new look from a shifting standards perspective. *Social Psychology of Education*, *20*, 471–490. doi:10.1007/s11218-017-9384-z

Jungbluth, P. (1993). Pygmalion and the effectiveness of "Black" schools: Teachers' stereotypes and hidden goal differentiations towards ethnic minorities. *Tijdschrift voor Onderwijsresearch*, *18*(2), 99–110.

Jussim, L., & Eccles, J. (1992). Teacher expectations: II. Construction and reflection of student achievement. *Journal of Personality and Social Psychology*, *63*, 947–961. doi:10.1037/0022-3514.63.6.947

Jussim, L., Eccles, J., & Madon, S. (1996). Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self-fulfilling prophecy. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 28, pp. 281–388). San Diego, CA: Academic Press.

Jussim, L., Robustelli, S. L., & Cain, T. R. (2009). Teacher expectations and self-fulfilling prophecies. In K. R. Wentzel, & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 349–379). New York, NY: Routledge.

Karing, C., Matthäi, J., & Artelt, C. (2011). Genauigkeit von Lehrerurteilen über die Lesekompetenz ihrer Schülerinnen und Schüler in der Sekundarstufe I – Eine Frage der Spezifität? [Lower secondary school teacher judgement accuracy of students' reading competence – A matter of specificity?]. *Zeitschrift für Pädagogische Psychologie*, *25*, 159–172. doi:10.1024/1010-0652/a000041

Kristen, C., Stanat, P., Kogan, I., Lorenz, G., Gentrup, S., & Rahmann, S. (2018). *Kompetenzerwerb und Lernvoraussetzungen (KuL)* [Competence acquisition and learning preconditions] (Version 1) [Data set]. Berlin: IQB – Institut zur Qualitätsentwicklung im Bildungswesen. doi:10.5159/IQB_KuL_v1

Kuhl, P., & Hannover, B. (2012). Differenzielle Benotungen von Mädchen und Jungen: Der Einfluss der von der Lehrkraft eingeschätzten Kompetenz zum selbstgesteuerten Lernen [Different grading for girls versus boys: Examining the impact of students' ability for self-regulated learning as perceived by their teachers]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *44*, 153–162. doi:10.1026/0049-8637/a000066

Latsch, M., & Hannover, B. (2014). Smart girls, dumb boys!? How the discourse on "failing boys" impacts performances and motivational goal orientation in German school students. *Social Psychology*, *45*, 112–126. doi:10.1027/1864-9335/a000167

Lorenz, G. (2018). *Selbsterfüllende Prophezeiungen in der Schule: Leistungserwartungen von Lehrkräften und Kompetenzen von Kindern mit Zuwanderungshintergrund* [Self-fulfilling prophecies in school: Teacher expectations and competences of immigrant children]. Wiesbaden: VS Verlag für Sozialwissenschaften.

Lorenz, G., Gentrup, S., Kristen, C., Stanat, P., & Kogan, I. (2016). Stereotype bei Lehrkräften? Eine Untersuchung systematisch verzerrter Lehrererwartungen [Stereotypes among teachers? A study of systematic bias in teacher expectations]. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *68*, 89–111. doi:10.1007/s11577-015-0352-3

Ludwig, P. H. (2007). Pygmalion zwischen Venus und Mars – Geschlechtsunterschiede in schulischen Lernleistungen durch Selbsterfüllung von Erwartungen [Pygmalion between Venus and Mars – Gender differences in academic achievements because of self-fulfilling expectations]. In P. H. Ludwig & H. Ludwig (Eds.), *Erwartungen in himmelblau und rosarot: Effekte, Determinanten und Konsequenzen von Geschlechterdifferenzen in der Schule* (pp. 17–59). Weinheim: Juventa Verlag.

Madon, S., Jussim, L., & Eccles, J. (1997). In search of the powerful self-fulfilling prophecy. *Journal of Personality and Social Psychology*, *72*, 791–809. doi:10.1037/0022-3514.72.4.791

McElvany, N., Kessels, U., Schwabe, F., & Kasper, D. (2017). Geschlecht und Lesekompetenz [Gender and reading achievement]. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, … R. Valtin (Eds.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (pp. 177–194). Münster: Waxmann.

McKown, C., & Weinstein, R. S. (2002). Modeling the role of child ethnicity and gender in children's differential response to teacher expectations. *Journal of Applied Social Psychology*, *32*, 159–184. doi:10.1111/j.1559-1816.2002.tb01425.x

McKown, C., & Weinstein, R. S. (2008). Teacher expectations, classroom context, and the achievement gap. *Journal of School Psychology*, *46*, 235–261. doi:10.1016/j.jsp.2007.05.001

Mengering, F. (2005). Bärenstark – Empirische Ergebnisse der Berliner Sprachstandserhebung an Kindern im Vorschulalter ["Bärenstark" – Empirical results of an investigation into the language competencies of preschool children in Berlin]. *Zeitschrift für Erziehungswissenschaft*, *8*, 241–262. doi:10.1007/s11618-005-0136-z

Merton, R. K. (1948). The self-fulfilling prophecy. *The Antioch Review*, *8*, 193–210. doi:10.2307/4609267

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). TIMSS 2015 International results in mathematics. Fourth grade mathematics. Exhibit 1.10. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2015/international-results/

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). *PIRLS 2016: International results in reading*. Chestnut Hill, IL: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College International Association for the Evaluation of Educational Achievement (IEA).

Organisation for Economic Co-operation and Development. (2016). *PISA 2015 results (Volume I): Excellence and equity in education*. Paris: Author. doi:10.1787/9789264266490-en

Pansu, P., Régner, I., Max, S., Colé, P., Nezlek, J. B., & Huguet, P. (2016). A burden for the boys: Evidence of stereotype threat in boys' reading performance. *Journal of Experimental Social Psychology*, *65*, 26–30. doi:10.1016/j.jesp.2016.02.008

Poloczek, S., Greb, K., & Lipowsky, F. (2009). Schulisches Selbstkonzept (Zusatzerhebung) [Academic self-concept (additional survey)]. In F. Lipowsky, G. Faust, & K. Greb (Eds.), *Dokumentation der Erhebungsinstrumente des Projekts "Persönlichkeits- und Lernentwicklung von Grundschülern" (PERLE) - Teil 1* (pp. 67–70). Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung.

Rauer, W., & Schuck, K. D. (2004). *FEESS 1-2: Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrungen von Grundschulkindern erster und zweiter Klassen* [FEESS 1-2: Questionnaire for emotional and social school experiences of first- and second-grade students]. Göttingen: Beltz Test.

Ready, D. D., & Wright, D. L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: The role of child background and classroom context. *American Educational Research Journal*, *48*, 335–360. doi:10.3102/0002831210374874

Robinson, J. P., & Lubienski, S. T. (2011). The development of gender achievement gaps in mathematics and reading during elementary and middle school: Examining direct cognitive assessments and teacher ratings. *American Educational Research Journal*, *48*, 268–302. doi:10.3102/0002831210372249

Robinson-Cimpian, J. P., Lubienski, S. T., Ganley, C. M., & Copur-Gencturk, Y. (2014). Teachers' perceptions of students' mathematics proficiency may exacerbate early gender gaps in achievement. *Developmental Psychology*, *50*, 1262–1281. doi:10.1037/a0035073

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*, *3*(1), 16–20. doi:10.1007/BF02322211

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: John Wiley & Sons.

Schrader, F.-W., & Helmke, A. (1990). Lassen sich Lehrer bei der Leistungsbeurteilung von sachfremden Gesichtspunkten leiten? Eine Untersuchung zu Determinanten diagnostischer Lehrerurteile [Are teachers influenced by extrinsic factors when evaluating scholastic performance? A study on the determinants of teachers' judgments]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *22*, 312–324.

Stahl, N. (2007). Schülerwahrnehmung und -beurteilung durch Lehrkräfte [Student perception and assessment by teachers]. In H. Ditton (Ed.), *Kompetenzaufbau und Laufbahnen im Schulsystem: Ergebnisse einer Längsschnittuntersuchung an Grundschulen* (pp. 171–198). Münster: Waxmann.

StataCorp LLC. (1985–2015). Stata 14.2 [Computer software]. College Station, TX: Author.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*, 797–811. doi:10.1037/0022-3514.69.5.797

Steffens, M. C., & Jelenec, P. (2011). Separating implicit gender stereotypes regarding math and language: Implicit ability stereotypes are self-serving for boys and men, but not for girls and women. *Sex Roles*, *64*, 324–335. doi:10.1007/s11199-010-9924-x

Tiedemann, J. (2002). Teachers' gender stereotypes as determinants of teacher perceptions in elementary school mathematics. *Educational Studies in Mathematics*, *50*, 49–62. doi:10.1023/A:102051810

Timmermans, A. C., De Boer, H., & Van der Werf, M. P. C. (2016). An investigation of the relationship between teachers' expectations and teachers' perceptions of student attributes. *Social Psychology of Education*, *19*, 217–240. doi:10.1007/s11218-015-9326-6

Trautner, H. M. (2008). Entwicklung der Geschlechtsidentität [Development of gender identity]. In R. Oerter & L. Montada (Eds.), *Entwicklungspsychologie* (pp. 625–651). Weinheim: Beltz.

Van Buuren, S., Brand, J. P. L., Groothuis-Oudshoorn, C. G. M., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Satistical Computation and Simulation*, *76*, 1049–1064. doi:10.1080/10629360600810434

Van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R. W. (2010). The implicit prejudiced attitudes of teachers: Relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal*, *47*, 497–527. doi:10.3102/0002831209353594
Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology*, *39*, 456–467. doi:10.1016/S0022-1031(03)00019-2
Weiß, R. H., & Osterland, J. (1997). *Grundintelligenztest Skala 1: CFT 1. Handanweisung* [Culture Fair Intelligence Test 1: CFT1. Manual] (5th rev. ed.). Göttingen: Hogrefe.
Wendt, H., Steinmayr, R., & Kasper, D. (2016). Geschlechterunterschiede in mathematischen und naturwissenschaftlichen Kompetenzen [Gender differences in mathematics and science achievement]. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert, & D. Kasper (Eds.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (pp. 257–298). Münster: Waxmann.

Appendix 1. Results of independent *t* tests testing the mean values of teacher expectation bias against zero

Variable	*n*	*M*	*SD*	*t*	*df*	*p*
German Language						
Negative Bias						
Total Sample	368	−0.81	0.44	−35.30	367	< .001
Girls	150	−0.74	0.37	−24.92	149	< .001
Boys	218	−0.86	0.48	−26.28	217	< .001
Positive Bias						
Total Sample	390	0.78	0.38	40.42	389	< .001
Girls	207	0.85	0.41	29.76	206	< .001
Boys	183	0.70	0.33	28.83	182	< .001
Mathematics						
Negative Bias						
Total Sample	369	−0.68	0.41	−32.22	368	< .001
Girls	193	−0.68	0.39	−24.42	192	< .001
Boys	176	−0.68	0.43	−21.19	175	< .001
Positive Bias						
Total Sample	369	0.69	0.37	35.29	368	< .001
Girls	170	0.71	0.40	23.09	169	< .001
Boys	199	0.67	0.35	26.93	198	< .001

Appendix 2. Gender-specific teacher expectancy effects in reading

	Model 1		Model 2	
Variable	*b* (*SE*)	*β*	*b* (*SE*)	*β*
Student Gender				
Female	−0.49 (2.04)	−.01	−0.17 (2.45)	−.01
Student Beginning-of-Year Achievement				
Phonological Awareness	1.85 (0.28)	.23***	1.87 (0.28)	.23***
Reading	0.04 (0.04)	.03	0.04 (0.04)	.03
Student Cognitive Abilities				
Deductive Reasoning	1.96 (0.41)	.15***	1.94 (0.42)	.15***
Working Memory	1.02 (0.56)	.05	1.01 (0.56)	.05

(*Continued*)

Continued.

Variable	Model 1 b (SE)	Model 1 β	Model 2 b (SE)	Model 2 β
Student Background				
Student Age	−0.76 (3.66)	−.01	−0.84 (3.64)	−.01
HISEI	0.18 (0.07)	.09*	0.18 (0.07)	.09*
Abitur	0.97 (2.88)	.03	0.98 (2.89)	.03
Immigrant Status	−1.38 (2.67)	−.04	−1.38 (2.68)	−.04
Teacher Perceptions				
Student Motivation and Work Habits	7.31 (1.68)	.17***	7.40 (1.67)	.17***
Teacher Expectation Bias				
German Language	3.93 (2.13)	.08	5.80 (3.04)	.12
German Language Squared	−0.58 (1.40)	−.01	0.26 (1.99)	.01
Female X German Language			−3.66 (3.22)	−.08
Female X German Language Squared			−0.55 (2.57)	−.01
Intercept	35.09 (25.19)		35.06 (25.04)	
R^2	26.04%		26.16%	
Cohen's f^2			0.00	

Note: Results of linear regression models predicting student achievement at the end of first grade. HISEI = Highest International Socio-Economic Index of Occupational Status. Cohen's f^2 = effect size of change in R^2 compared to the prior model. N = 1,025. *p < .05. **p < .01. ***p < .001.

Appendix 3. Gender-specific responses to negative and positive teacher expectation bias in reading

Variable	b (SE)	95% CI for OR: Lower	OR	Upper
Student Gender				
Female	−0.42 (0.29)	0.37	0.66	1.15
Teacher Expectation Bias				
Positive Bias	0.05 (0.34)	0.55	1.05	2.03
Female X Positive Bias	0.49 (0.34)	0.84	1.62	3.14
Student Background				
Student Age	−0.08 (0.31)	0.51	0.92	1.68
HISEI	0.01 (0.01)	1.00	1.01	1.02
Abitur	0.03 (0.22)	0.66	1.03	1.60
Immigrant Status	−0.15 (0.21)	0.57	0.86	1.31
Intercept	0.53 (2.10)			
N	530			

Note. Results of logistic regression models predicting confirming versus disconfirming student achievement development. HISEI = Highest International Socio-Economic Index of Occupational Status. *p < .05. **p < .01. ***p < .001.

Appendix 4. Gender-specific teacher expectancy effects in mathematics

Variable	Model 1 *b* (*SE*)	Model 1 *β*	Model 2 *b* (*SE*)	Model 2 *β*
Student Gender				
Female	−1.67 (0.36)	−.24***	−1.69 (0.44)	−.24***
Student Beginning-of-Year Achievement				
Mathematics	0.55 (0.03)	.63***	0.55 (0.03)	.63***
Student Cognitive Abilities				
Deductive Reasoning	0.27 (0.06)	.11***	0.27 (0.06)	.11***
Working Memory	0.08 (0.09)	.02	0.08 (0.09)	.02
Student Background				
Student Age	−0.57 (0.47)	−.03	−0.57 (0.47)	−.03
HISEI	0.02 (0.01)	.05	0.02 (0.01)	.05
Abitur	0.44 (0.34)	.06	0.45 (0.35)	.06
Immigrant Status	−0.18 (0.36)	−.03	−0.18 (0.36)	−.03
Teacher Perception				
Student Motivation and Work Habits	0.16 (0.28)	.02	0.17 (0.28)	.02
Teacher Expectation Bias				
Mathematics	0.76 (0.28)	.07**	0.92 (0.33)	.09**
Mathematics Squared	0.09 (0.19)	.01	0.10 (0.24)	.01
Female X Mathematics			−0.33 (0.44)	−.03
Female X Mathematics Squared			0.02 (0.38)	.00
Intercept	24.67*** (3.16)		24.62*** (3.15)	
R^2	57.49%		57.52%	
Cohen's f^2			0.00	

Note: Results of linear regression models predicting student achievement at the end of first grade. HISEI = Highest International Socio-Economic Index of Occupational Status. Cohen's f^2 = effect size of change in R^2 compared to the prior model. $N = 1{,}003$. *$p < .05$. **$p < .01$. ***$p < .001$.

Appendix 5. Gender-specific responses to negative and positive teacher expectation bias in mathematics

		95% CI for *OR*		
Variable	*b* (*SE*)	*Lower*	*OR*	*Upper*
Student Gender				
Female	0.61** (0.22)	1.20	1.84	2.83
Teacher Expectation Bias				
Positive Bias	0.52 (0.33)	0.87	1.69	3.25
Female X Positive Bias	−1.44*** (0.41)	0.11	0.24	0.52
Student Background				
Student Age	0.17 (0.32)	0.63	1.19	2.24
HISEI	−0.00 (0.01)	0.99	1.00	1.01
Abitur	0.08 (0.23)	0.69	1.08	1.71
Immigrant Status	−0.15 (0.22)	0.56	0.86	1.34
Intercept	−0.94 (2.15)			
N	447			

Note: Results of logistic regression models predicting confirming versus disconfirming student achievement development. HISEI = Highest International Socio-Economic Index of Occupational Status. $^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.