

Self-fulfilling prophecies in the classroom: Teacher expectations, teacher feedback and student achievement



Sarah Gentrup^{a,*}, Georg Lorenz^b, Cornelia Kristen^c, Irena Kogan^d

^a Institute for Educational Science, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099, Berlin, Germany

^b Institute for Educational Quality Improvement (IQB), Humboldt-Universität zu Berlin, Unter den Linden 6, 10099, Berlin, Germany

^c Chair for Sociology, esp. Analysis of Social Structures, Otto-Friedrich-Universität Bamberg, Feldkirchenstraße 21, 96045, Bamberg, Germany

^d Mannheim Centre for European Social Research (MZES), Universität Mannheim, 68131, Mannheim, Germany

ARTICLE INFO

Keywords:

Self-fulfilling prophecy
Teacher expectations
Teacher feedback
Teacher-student interaction
Videography

ABSTRACT

This study investigated the link between teacher expectations and student learning, relying on longitudinal data from 64 classrooms and 1026 first-grade students in Germany. Further, based on a subsample of 19 classrooms with 354 students, we explored the mediating role of three characteristics of teacher feedback rated in video-recorded school lessons. The results showed that teacher expectations were inaccurate to some extent; that is, they did not entirely agree with students' current achievement, general cognitive abilities and motivations. In addition, this inaccuracy in teacher expectations significantly predicted students' end-of-year achievement, even after prior achievement, general cognitive abilities, motivation, and student background characteristics were considered. Specifically, inaccurately high teacher expectations were associated with greater achievement in reading and mathematics, whereas inaccurately low teacher expectations were associated with lower achievement in reading only. Furthermore, teacher feedback varied significantly with inaccurate teacher expectations but did not substantially mediate teacher expectancy effects.

1. Introduction

From the first day of school enrollment onward, teachers play a significant role in student learning (e.g., Dietrich, Dicke, Kracke, & Noack, 2015; Hattie, 2009; Roorda, Koomen, Spilt, & Oort, 2011). To promote student learning, it is crucial that teachers are aware of students' achievement, as well as their individual learning resources, as this knowledge is the basis for effective instructional decisions and enables teachers to provide sufficient support to individual students (e.g., Baumert & Kunter, 2013; Vogt & Rogalla, 2009). Such evaluations include not only perceptions of current student achievement but also expectations about students' learning and future achievement (Funder, 1995; Jussim, Robustelli, & Cain, 2009). At the same time, teachers' inaccurate achievement expectations can result in a self-fulfilling prophecy; that is, low expectations can hamper students' learning, whereas high expectations can foster students' learning and eventually lead to higher achievement gains. Rosenthal and Jacobson (1968), with their experiment "Pygmalion in the Classroom", were the first to provide evidence of a self-fulfilling prophecy (Merton, 1948) in the context of school. The experiment was controversial, with several scholars questioning different aspects of its results (Pellegrini & Hicks, 1972; Snow,

1969; Thorndike, 1968). Today, researchers generally agree that teacher expectancy effects exist (for overviews, see Jussim et al., 2009; Wang, Rubie-Davies, & Meissel, 2018). Teacher achievement expectations can thus affect child development from the first school days onward, as well as affecting later educational achievements and eventual outcomes.

Nonetheless, the question of how teacher expectancy effects emerge has not been conclusively answered. Most authors agree with the assumption that self-fulfilling prophecies in the classroom follow a sequence of three major steps (Jussim et al., 2009): (1) teachers form inaccurate expectations; (2) these expectations lead teachers to treat higher- and lower-expectancy students differently; and (3) students react to this differential teacher treatment in such a manner that confirms the initial teacher expectations, hence resulting in greater achievement gains for higher-expectancy students and lower achievement gains for lower-expectancy students. Until now, however, very few studies have provided empirical evidence of this three-step formation process; specifically, for the mediation of teacher expectancy effects by differential teacher behavior. Past studies have examined only one or two of the steps. Whereas there has been much research on accuracy in teacher expectations (for an overview, see Jussim et al.,

* Corresponding author.

E-mail address: sarah.gentrup@hu-berlin.de (S. Gentrup).

2009; for a meta-analysis, see Südkamp, Kaiser, & Möller, 2012) and the effects of teacher expectation inaccuracy on student achievement (for overviews, see Jussim, Eccles, & Madon, 1996; Jussim et al., 2009), relatively few studies have addressed the ways in which teachers communicate their expectations through their behavior, and even fewer studies have investigated whether such differential teacher behavior actually mediates teacher expectancy effects. Thus, there is a dearth of research on the effects of differential teacher behaviors triggered by teacher expectations on student outcomes. In the early years of expectancy research, four dimensions of teacher behavior were discussed as mediators of teacher expectancy effects: input (e.g., amount and difficulty of learning material provided), output (e.g., calling on a student), feedback (e.g., valence and elaborateness of feedback given) and climate (e.g., warmth and respectfulness in teacher-student interaction) (Rosenthal, 1973). Although empirical research has provided support for the relevance of each of these four dimensions (Harris & Rosenthal, 1985), evidence regarding the feedback dimension appears to be inconclusive. Findings on the relationships between different indicators of teacher feedback and teacher expectations, on the one hand, and student learning, on the other hand, have varied significantly, resulting in contradictory conclusions (Harris & Rosenthal, 1985). However, since there have been few newer empirical studies on the mediation of teacher expectancy effects, the recent research still relies on the four-factor model.

The present paper was based on a unique longitudinal data set that enabled us to examine the relationships between initial teacher expectations and later teacher feedback, as well as subsequent student achievement immediately at the beginning of students' school careers in Germany. Our study had two main goals. (1) By empirically tracing all three steps of the process potentially leading to the emergence of self-fulfilling prophecies in school, we sought to clarify whether associations between teacher expectations (as measured at the beginning of first grade) and student achievement (as measured at the end of first grade) are mediated by differences in teacher behavior, specifically teacher feedback. Thus, we directly investigated whether the three-step process happens as assumed. (2) Further, our study aimed to contribute to resolving open questions regarding the role of teacher feedback as a mediator of teacher expectancy effects. The inconclusive results in Harris and Rosenthal's meta-analysis show that this dimension of teacher behavior requires further investigation.

2. Theory and evidence

2.1. Inaccurate teacher expectations

Teacher expectations can be viewed as predictions of future student achievement (e.g., Jussim, 1986; Ready & Chu, 2015). Many researchers have only considered students' current achievements to be accurate bases of teacher expectations (e.g., Hinnant, O'Brien, & Ghazarian, 2009; McKown & Weinstein, 2002; Ready & Chu, 2015; Rubie-Davies & Peterson, 2016). From a pedagogical point of view, however, further student characteristics, such as general cognitive ability and motivation (in contrast to other student characteristics, such as gender or socioeconomic status, which should be irrelevant in a meritocratic school system) can also serve as valid predictors of teacher expectations (cf. de Boer, Bosker, & van der Werf, 2010). Based on these considerations and in line with research by Becker (2013) and de Boer et al. (2010), we characterize teacher expectations as accurate if they concur with actual students' achievement, general cognitive abilities and achievement motivation. Deviations in teacher expectations that exist beyond these learning-related student characteristics are hence referred to as teacher expectation inaccuracy.

Prior research has shown that teacher expectations for students' achievements are accurate to a substantial degree (Jussim et al., 2009). In meta-analyses, the shared variance between teacher judgments of current student achievement and students' actual achievements

amounted to approximately 40% (Hoge & Coladarci, 1989; Südkamp et al., 2012). The explained variance in teacher expectations was somewhat higher when general cognitive abilities and motivation were considered as valid predictors of teacher expectations (Becker, 2013; de Boer et al., 2010). Nevertheless, a significant portion of the variance in teacher expectations remained unexplained by students' learning-related characteristics. This variance reflects inaccuracy and has the potential to initiate self-fulfilling prophecies.

Furthermore, studies have shown that inaccuracy in teacher expectations does not occur randomly but is systematic for different groups of students. For example, negative bias in teacher expectations has been found for students from socially disadvantaged families (e.g., de Boer et al., 2010; Lorenz, Gentrup, Kristen, Stanat, & Kogan, 2016; Ready & Wright, 2011; Timmermans, de Boer, & van der Werf, 2016; Timmermans, Kuyper, & van der Werf, 2015; van Matre, Valentine, & Cooper, 2000), for ethnic minority students (e.g., Holder & Kessels, 2017; Lorenz et al., 2016; McKown & Weinstein, 2008; Meissel, Meyer, Yao, & Rubie-Davies, 2017; Morris, 2005; Ready & Wright, 2011), for boys and girls in gender-untypical domains (e.g., Hinnant et al., 2009; Holder & Kessels, 2017; Lorenz et al., 2016; Meissel et al., 2017; Ready & Wright, 2011; Riegle-Crumb & Humphries, 2012; Rubie-Davies & Peterson, 2016) as well as for students with special educational needs or learning disability statuses (e.g., Hurwitz, Elliott, & Braden, 2007; Jenkins & Demaray, 2016; Shifrer, 2013, 2016). Teacher expectancy effects based on such systematically biased expectations have the potential to contribute to educational inequalities (e.g., de Boer et al., 2010; Gentrup & Rjosk, 2018; Muntoni & Retelsdorf, 2018). However, as the main focus of the current study is the mediation of teacher expectancy effects via teacher feedback, we do not investigate group-specific bias but rather general teacher expectation inaccuracy.

2.2. Teacher expectations and differential teacher behavior

Most studies examining teacher behavior in the context of teacher expectancy effects are from the 1970s and 1980s (for a meta-analysis, see Harris & Rosenthal, 1985; for a review, see Jussim, 1986). In the last 30 years, very few studies investigated this issue (e.g., Chen, Thompson, Kromrey, & Chang, 2011; Ready & Chu, 2015; for a recent review, see Wang et al., 2018). Therefore, the contemporary expectancy research still relies on the four-factor model proposed by Rosenthal (1973), which was empirically evaluated in the meta-analysis by Harris and Rosenthal (1985). This model assumes that teachers' expectations affect four dimensions of their behavior (cf. Jussim et al., 2009). (1) Teachers might differ in the *input* they provide. For example, teachers might explain issues in a less complex manner to lower-expectancy students than to higher-expectancy students. (2) According to their expectations, teachers might provide different opportunities for students to produce *output*. For example, teachers might call on lower-expectancy students less frequently than they call on higher-expectancy students. (3) Teacher expectations might influence teacher *feedback*, which could be less positive and less constructive for lower-expectancy students than for higher-expectancy students. (4) Finally, the *climate* of teacher-student interactions might be less warm and respectful for lower-expectancy students than for higher-expectancy students. The results of the meta-analysis (Harris & Rosenthal, 1985) supported the relevance of each of the four proposed dimensions, with *climate* and *input* showing the strongest relationships with both teacher expectations and student outcomes. The evidence for *feedback*, however, remained inconclusive. The relationships of diverse indicators of teacher feedback with teacher expectations and student achievement varied significantly and showed, in part, contradicting results. The association between teacher expectations and *feedback* varied, depending on the indicators used, between $r = -.05$ and $r = 0.36$, showing in summary a small effect size ($r = 0.13$). The observed relationship between indicators of *feedback* and student achievement even ranged from $r = -.23$ to $r = 0.12$. Overall, the link between feedback and student achievement amounted

to $r = 0.07$. Thus, it appears that diverse indicators of feedback, collapsed into one feedback dimension, operate in opposite directions – with the consequence of an average effect size of approximately zero.

Despite providing important insights into the relationship between teacher expectations and teacher behavior, the meta-analysis also had its limitations. First, the authors did not ensure that the studies included in their meta-analysis controlled for actual student achievement. Therefore, it was not possible to determine whether the differential teacher treatment resulted from accurate teacher expectations or from teacher expectation inaccuracy. Second, only a few of the studies included in the meta-analysis investigated both the association between teacher expectations and teacher behavior and the association between teacher behavior and student outcomes. The meta-analysis thus did not directly address mediation effects (cf. Chow, 1987).

Newer studies on the relationships between teacher expectations and teacher behavior are few, and they largely suffer from the same limitations. Chen et al. (2011) investigated teacher expectancy effects on student self-concept in elementary schools in Taiwan and considered students' perceptions of oral feedback given by their teachers. The results showed that lower-expectancy students perceived receiving less positive and more negative oral feedback from their teachers than higher-expectancy classmates. Furthermore, the perceived feedback was significantly related to students' self-concept. However, this study also did not account for actual student achievement. Furthermore, it leaves open whether teacher feedback actually mediates the link between teacher expectations and students' self-concept.

Other newer studies examined teacher behavior in relation to teacher expectations at the classroom level (e.g., Rubie-Davies, Hattie, Townsend, & Hamilton, 2007; Rubie-Davies & Peterson, 2011). Although these studies supported the concept that teachers with high expectations for all of their students are more effective in teaching (e.g., provide more feedback and use more higher-order questions; Rubie-Davies, 2007), they also left open the question of whether the observed differences in teacher behavior actually mediate teacher expectancy effects.

Two newer studies directly investigated the mediation of teacher expectancy effects by teacher behavior. First, Urhahne (2015) analyzed the effects of teacher expectations on students' motivations and emotions. Differential teacher behaviors measured through student perceptions were found to partially mediate the link between teacher expectations and students' motivations and emotions. Second, Ready and Chu (2015) examined whether teachers' practice of ability grouping (which can be viewed as an indicator of *input*) mediates the link between individual teacher expectations and student achievement. Contradicting the findings by Harris and Rosenthal (1985) on the input dimension of teacher behaviors, ability grouping was found to be only a weak mediator of teacher expectancy effects (Ready & Chu, 2015). This limited and inconclusive evidence on the mediation of teacher expectancy effects via differential teacher behavior calls for research based on data that are better suited to addressing the theoretical arguments. The current study focuses on teacher feedback as a possible mediator of teacher expectancy effects.

2.2.1. Teacher expectations and teacher feedback

Teachers' expectations may shape their feedback practice on the basis of mental schemata. Based on their teaching experience, teachers can have internalized schemata of *high-achieving* and *low-achieving students*, that is, mental representations about typical characteristics and behaviors of high-achieving and low-achieving students, as well as perceptions about the appropriate teaching of these students (Fazio & Olson, 2014; Pendry, 2015). As an interview study of approximately 300 student teachers revealed, several positive attitudes are associated with high-achieving students. Such students are seen as interested and motivated, as exerting effort for their learning, and as showing disciplined behavior. Low-achieving students, in contrast, are perceived as undisciplined, uninterested, unmotivated and unintelligent (Schuchardt

& Dunkake, 2014). Teachers can interpret student behavior and achievement based on these mental representations of high-achieving and low-achieving students, and such interpretations can result in predetermined (patterns of) reactions (Dijksterhuis & Bargh, 2001).

According to the results of the study by Schuchardt and Dunkake (2014), the schema of a low-achieving student is connected, for instance, to the assumption of an undisciplined child. Teachers might thus be more sensitive to the misbehavior of students whom they categorize as low-achievers (that is, low-expectancy students) and notice misbehavior more often for these students than for students whom they categorize as high-achievers (that is, high-expectancy students). Consequently, teachers might more often give children for whom they have lower expectations feedback related to their behavior in the classroom, instead of feedback related to their performance.

As the schema of a high-achieving student also includes favorable student characteristics in general (cf. Schuchardt & Dunkake, 2014), teachers might interpret the achievement output of high-expectancy students to be more positive than the similar outputs of low-expectancy students (*confirmation bias*; Nickerson, 1998). That is, the same student achievement might trigger more or less positive performance feedback depending on the schema activated. Accordingly, teachers might more often praise students for whom they have higher expectations, whereas negative performance feedback might be more frequent for lower-expectancy students, independent of the students' actual achievements. Eventually, the valence of performance feedback might be more positive for higher-expectancy students than for lower-expectancy students.

As studies relying on attribution theory have indicated (Johnson, Feigenbaum, & Weiby, 1964), teachers also feel stronger self-efficacy in teaching high-achieving students. Therefore, when interacting with high-expectancy students, teachers' performance feedback might be more elaborate. For example, negative performance feedback might be more often accompanied by further elaboration and tips on how to proceed for higher-expectancy students than for lower-expectancy students. Teachers might also explain in greater detail how students have misunderstood the learning material or which types of errors they have made.

2.2.2. Role of teacher feedback in student learning

Educational research has supported the importance of teacher feedback for student learning (e.g., Bohn, Roehrig, & Pressley, 2004; Dean, Hubbell, Pitler, & Stone, 2012). In fact, teacher feedback has been empirically identified as being among the most important instructional practices for improving student learning. In a meta-analysis, high-quality feedback had an average effect size on student achievement of $d = 0.73$ and was among the top ten investigated instructional practices (Hattie, 2009). Additionally, when children were asked about possible sources of information regarding their own levels of achievement in school, they primarily referred to teacher feedback (Weinstein, 1983).

Feedback provides useful information about how well a student is performing. Frequent and informative teacher feedback therefore helps students to overcome mistakes and improve skills. In contrast, if feedback is less frequently provided or less informative, students might not be aware that they have not fully mastered the material and that they must improve their skills (c.f. Jussim et al., 2009). The literature has further emphasized that individualized teaching styles, in particular, require high-quality feedback. Such feedback must be specific, task-oriented and related to students' learning goals; it must evaluate how well students are performing to reach those goals; and it must provide information to students on how to proceed (Hattie & Timperley, 2007). Therefore, feedback related to students' performance should promote students' learning in reading and mathematics better than feedback related to students' behavior in class. To strengthen students' perceived self-efficacy, a higher intraindividual ratio of positive and therefore affirmative performance feedback should be pedagogically beneficial (cf. Bandura, 1994). Additionally, especially in the case of negative

Table 1
Research questions and hypotheses of the current study.

| Research questions | Hypotheses |
|--|---|
| (1) Are teacher expectations of students' achievement inaccurate and, if so, to what extent? | (1) The variance in teachers' expectations cannot be fully explained by actual student skills, cognitive abilities, and motivation. |
| (2) Are inaccurate teacher expectations reflected in teachers' feedback practices? | (2) Higher values on the variable of teacher expectation inaccuracy are associated with: (a) more performance-related feedback (compared to feedback on student behavior); (b) more positive performance feedback; and (c) greater elaborateness of negative performance feedback. |
| (3) Do inaccurate teacher expectations predict student learning in reading and mathematics beyond other student characteristics that affect students' learning progress? | (3) In reading and mathematics: (a) inaccurately high expectations are related to higher achievement gains; and (b) inaccurately low expectations are related to lower achievement gains over time. |
| (4) Do teachers' feedback practices mediate teacher expectancy effects? | (4) Teacher feedback partially mediates teacher expectancy effects. |

performance feedback, it is important that teachers provide detailed information about learning goals and how to proceed in learning.

2.3. Teacher expectancy effects on students' achievement

The effects of teacher expectations on student achievement have been extensively studied (for overviews, see Jussim et al., 1996; Jussim & Harber, 2005; Jussim et al., 2009). The bulk of research comes from the U.S. (e.g., Gill & Reynolds, 1999; Hinnant et al., 2009; Ma, 2001; McKown & Weinstein, 2008; Ready & Chu, 2015; Rubie-Davies et al., 2014), and a few studies were conducted in Europe (e.g., Germany: Friedrich, Flunger, Nagengast, Jonkmann, & Trautwein, 2015; Lorenz, 2018; Stahl, 2007; Netherlands: de Boer et al., 2010; Jungbluth, 1993, 1994). Despite significant methodological differences, most studies have found support for teacher expectancy effects on student achievement (cf. Wang et al., 2018). The effects in field studies varied between $r = 0.10$ and $r = 0.20$; in terms of the effect size d , the average effect was $d = 0.30$ (Jussim et al., 2009).

However, most studies on teacher expectancy effects examined students at later stages of their educational trajectories, that is, at the later grade levels of primary school or secondary school (e.g., Archambault, Janosz, & Chouinard, 2012; de Boer & van der Werf, 2015; Friedrich et al., 2015; Peterson, Rubie-Davies, Osborne, & Sibley, 2016; Zhou & Urhahne, 2013). Expectancy research focusing on students' early school career is comparatively sparse. Recently, Schenke, Nguyen, Watts, Sarama, and Clements (2017) analyzed the link between class-level teacher expectations and 4-year-old students' mathematical achievement. Ready and Chu (2015) investigated whether students' literacy achievement in kindergarten was affected by teacher expectations. The results of both studies suggest that teacher expectancy effects can manifest in such early years. Furthermore, research supports that teacher expectancy effects are stronger in new situations such as after school enrollment or school transitions (cf. Jussim et al., 1996; Jussim et al., 2009). For example, in the study by Kuklinski and Weinstein (2001), teacher expectancy effects were stronger in first grade than in the later grades of elementary school. Similarly, for the period of five years after the transition into secondary school, de Boer et al. (2010) found that the effects of teacher expectations were strongest in the first year, whereas they decreased somewhat in the second year and remained stable afterwards.

A common criticism of studies of expectancy effects is that teacher expectations might have predicted student outcomes simply because they were accurate representations of students' prior achievements (Jussim & Harber, 2005). The main challenge of measuring expectancy effects is thus to disentangle the statistical associations of student outcomes with accurate teacher expectations, on the one hand, from their associations with inaccurate teacher expectations, on the other hand. Only the latter association indicates teacher expectancy effects (de Boer et al., 2010). As Wang et al. (2018) outlined, approximately 40% of the studies on teacher expectancy effects conducted in the last 30 years did not consider actual student characteristics in their analyses and thus did not adequately account for this challenge.

In addition, most existing studies measured teacher expectations not at the beginning of the teacher-student interaction but after months or even years of such interaction (e.g., Friedrich et al., 2015; Hinnant et al., 2009; Peterson et al., 2016; Ready & Chu, 2015). In this situation, it is not possible to disentangle whether teacher expectations concur with actual student characteristics because they are accurate evaluations of these characteristics or because of preceding self-fulfilling prophecies. Against this background and the observation that approximately 40% of the studies did not account for actual student achievement (Wang et al., 2018), the validity of the existing results might be limited.

3. The current study

In summary, although evidence for the existence of teacher expectancy effects on student achievement abounds, only a few studies have sought to clarify how these effects emerge. Accordingly, there has been little research on teacher behaviors mediating the link between inaccurate teacher expectations and student achievement. In particular, the results for teacher behavior in terms of feedback have been inconclusive. Furthermore, methodological limitations might reduce the validity of research on expectancy effects. These limitations include a failure to consider baseline student achievements and the problem of measuring teacher expectations after prolonged periods of teacher-student contact.

With the aim of overcoming these shortcomings, the present study examined the associations of teacher expectations with teacher feedback and student learning, covering all three steps of the process potentially leading to the emergence of a self-fulfilling prophecy. We used data from a unique longitudinal study in Germany explicitly designed to investigate self-fulfilling prophecies in schools. Data collection started immediately after the beginning of the first school year, when preceding teacher-student interaction was minimal. We analyzed the processes underlying teacher expectancy effects in two domains: language and mathematics. More specifically, we studied teachers' feedback practices and explored whether they differed depending on teachers' expectations. We also investigated the extent to which feedback mediates teacher expectancy effects on student achievement in reading and mathematics. Table 1 contains the specific research questions that we addressed, as well as the corresponding hypotheses.

4. Methods

4.1. Databases

The study was based on data from the research project *Kompetenzerwerb und Lernvoraussetzungen*¹ (KuL; English translation:

¹ This interdisciplinary research project was conducted under the leadership of Prof. Dr. Cornelia Kristen (Otto-Friedrich-Universität Bamberg), Prof. Dr. Irena Kogan (Universität Mannheim), and Prof. Dr. Petra Stanat (Humboldt-Universität zu Berlin).

Competence Acquisition and Learning Preconditions). In the 2013/2014 school year, data were collected in 39 primary schools in North Rhine-Westphalia, Germany. The total sample of the research project included 1065 first graders from 67 classes (Kristen et al., 2018a). We excluded from the sample classes in which the teachers changed during the school year, which left $N = 1026$ students from $N = 64$ classes with $N = 67$ class and subject teachers in $N = 38$ schools. This sample is referred to as the *main sample*. In the main sample, on average, 16 students per class participated in the study ($SD = 5.25$), which corresponds to a participation rate of 68% ($SD = 21\%$).

The teachers (94% female) were, on average, 42 years old ($SD = 8.80$) and had an average work experience of twelve years ($SD = 8.89$). The teachers were predominantly nonimmigrant, meaning that they and both of their parents were born in Germany (94%). At the time of school enrollment, the participating students (48% female) in the main sample were, on average, 6 years and 5 months old ($SD = 0.33$). Based on information from the parent questionnaire, 36% of the children came from immigrant families (at least one parent born in a country other than Germany). The average families' socioeconomic background, as indicated by the Highest International Socio-Economic Index of Occupational Status (HISEI; Ganzeboom, 2010), was $M = 52.51$ ($SD = 19.44$).

A subsample of $n = 354$ children from $n = 19$ classrooms in $n = 13$ schools participated in the optional video study, which was conducted in the middle of the school year (Kristen et al., 2018b). Despite the voluntary participation of teachers and students, the subsample of classrooms did not differ significantly from the main sample in important characteristics. The teachers involved in the video study did not differ substantially from the main sample in terms of gender (95% female), immigrant status (95% nonimmigrant), age ($M = 41.85$, $SD = 9.51$) or years of professional experience ($M = 12.11$, $SD = 8.60$). The students in the video subsample (47% female) were also similar to those in the main sample regarding their age ($M = 6.48$, $SD = 0.35$) and socioeconomic background ($M = 52.03$, $SD = 19.69$). Slightly more students in the video subsample than in the main sample came from immigrant families (38%). In addition, the participation rate per class was substantially higher in the subsample ($M = 86\%$, $SD = 9\%$). Descriptive statistics of further student variables in the main sample and subsample are displayed in Appendix A.

4.2. Survey design

The KuL study was a mini-panel that tracked first graders and their teachers in the school subjects of German language and mathematics throughout the first grade. The study received approval from the research ethics committee of *Universität Mannheim*. Data collection was performed in three waves. The first wave occurred immediately after the beginning of the school year and included standardized achievement tests and interviews with the students (September to November 2013, during the third to ninth school weeks), a questionnaire for teachers (dispatched by the research team in September to November 2013, during the third to seventh school weeks) and telephone interviews with the parents (conducted in October to December 2013). The second survey wave in the middle of the school year (February and March 2014) included video recordings of teacher-student interactions during lessons and further interviews with the students. At the end of the school year, the students were tested and interviewed again (May and June 2014), and the teachers completed another questionnaire (dispatched by the research team at the beginning of May 2014). All three study waves and the instruments were pretested in two separate schools in the year preceding the main study.

The resulting data have two important advantages. First, both teacher expectations and student achievement were measured immediately after school enrollment. Hence, measured student achievement should not yet have been influenced by teachers' expectations and behaviors, as the teacher-student interaction prior to the data collection

was minimal. Second, the video recordings of classroom interactions allowed us to examine teacher feedback practices and link them to the teachers' expectations, on the one hand, and to student achievement, on the other hand. The data set thus enabled us to analyze the mediating mechanisms of self-fulfilling prophecies in schools considering a range of indicators of both teacher expectations and student characteristics.

4.3. Instruments

4.3.1. Teacher expectations

Teachers rated each of the participating students in their classes on five items (5-point-scale), indicating their expectations for each child's achievement in the German language (three items; $\alpha = 0.94$) and mathematics (two items; $\alpha = 0.94$). Teachers were asked to compare the skill levels that they expected the children to acquire by the end of first grade to that of their classmates (e.g., "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"; see Appendix B). 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.

To investigate teacher expectation inaccuracy, we applied the residual approach proposed by Madon, Jussim, and Eccles (1997). The procedure is explained in section 4.4.2.

4.3.2. Student achievements and abilities

At the beginning and end of first grade, the students completed tasks of the subscales *phonological awareness* ($\alpha = 0.82$) and *reading* ($\alpha = 0.96$) from the *Fähigkeitsindikatoren Primarschule (FIPS)* computer-based assessment (German version of the *Performance Indicators in Primary Schools (PIPS)*; Bäuerlein et al., 2012) as measures of language skills and the subscale *mathematics* ($\alpha = 0.92$) as a measure of mathematical skills. Students also completed a deductive reasoning test (*CFT*; Weiß & Osterland, 1997; $\alpha = 0.78$) and the subscale *working memory* implemented in the FIPS assessment (Bäuerlein et al., 2012; $\alpha = 0.76$) at the beginning of the school year. The two scales captured students' general cognitive abilities.

4.3.3. Student motivation

In addition to students' achievement and general cognitive abilities, teachers might also rely on students' motivation when forming their expectations. In study wave 2, all participating children were interviewed regarding their enjoyment of learning (13 items; $\alpha = 0.78$) and the effort that they invest in learning (13 items; $\alpha = 0.70$). Both motivational traits were measured with an adapted version of a questionnaire by Rauer and Schuck (2004). For each item, the students indicated whether it applied to them (enjoyment of learning: e.g., "I like to learn at school"; effort: e.g., "I also try to solve very difficult tasks"). To increase differentiation, we used a 3-point scale (0 = *not true*, 1 = *partly true*, 2 = *completely true*) instead of the original dichotomous response scale (0 = *not true*, 1 = *true*). As indicated by mean scores greater than the central point of the response scales (see Appendix A), the first graders in this sample stated that they fully enjoyed learning and were exerting much effort in learning. This pattern is typical for young students (e.g., Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002; Spinath & Steinmayr, 2008).

4.3.4. Teacher feedback practice

In the middle of the school year, we video-recorded between two and four school lessons (approximately 45 min per lesson) in each class that participated in the video study. Usually, two lessons were recorded in each subject (i.e., German language and mathematics). We asked teachers to set up their lessons as usual. We also asked them to include phases in which they interact with the whole class as well as phases in which the students work on their own. As we scheduled the specific appointments for video-recording with each teacher, they knew

beforehand which lessons would be video-recorded. The videos were then coded by independent raters who had not been informed about any specific research questions or hypotheses of the study. The coding followed the method of event-sampling.

In a first step, the raters identified each time sequence in which the teacher directed his or her attention, an action or a statement toward one to three students at a time (coding of *interaction sequences*). Because the coding aimed at identifying differences in teacher behavior for individual students, time periods in which the teacher interacted with more than three students at a time (for example, with the whole class) were ignored. In a second step, each interaction sequence was chopped into subsequences (coding of *subsequences*) based on changes in the content of the interaction happening within the interaction sequence (e.g., feedback, elaborations, and instructions). In a final step, each subsequence was then rated in terms of its content (coding of *content*). Here, raters considered, i.e., a range of feedback codes (i.e., very positive/positive/negative/very negative performance/behavioral feedback) and elaboration codes (i.e., direct hint/prompt/supporting question).

The training of the raters was based on video recordings of school lessons from the pretest study. We conducted two training periods and assessed the raters' reliability after each period. For the first reliability check, each rater coded an exemplary video with predefined interaction sequences. Thus, the raters had to decide about the timing and content of the subsequences. The results were compared to a master coding performed by the first author of this study, who had also developed the coding guidelines. Estimates of the rater-master agreement are displayed in the upper half of [Appendix C](#). A second training period was conducted after the raters coded video recordings of 14 school lessons from the main study (21.5% of all video recordings). Part of this reliability check followed the same procedure as the first reliability check (i.e., video raters had to decide about the timing and content of subsequences within predefined interaction sequences). Additionally, the raters coded a second exemplary video in which not only the interaction sequences but also the timing of 168 subsequences were predefined. Here, the video raters had to focus exclusively on the content of the subsequences. Based on these two codings, we tested for three aspects of reliability, all indicating the average agreement of the raters' codings with the master coding (see the lower half of [Appendix C](#)). First, we split every interaction sequence coded by the video raters into 100-msec sections and calculated Cohen's kappa by comparing them with the same 100-msec sequences from the master coding. The resulting estimate indicates the content-related agreement between a rater's coding and the master coding, while considering differences in the coding of the timing of subsequences. Cohen's kappa varied between $\kappa = 0.62$ and $\kappa = 0.70$, indicating substantial agreement (see [Landis & Koch, 1977](#)). Second, we focused on the additional content-related rating of 168 subsequences that had been predefined with regard to their timing. The resulting kappa estimates amounted to values between $\kappa = 0.96$ and $\kappa = 0.97$, indicating high, almost perfect agreement. Finally, we investigated the correlations between the rater coding and master coding for each of the three teacher feedback variables used in the analyses (see the following paragraphs for more information). The correlations varied between $r = 0.70$ and $r = 0.97$. These scores indicate medium to high agreement.

The coding of the videos focused on different contents of the interactions that occurred between teachers and students. One type of content relevant for the present study was feedback, which was further categorized into performance feedback (e.g., "You read very well") and behavioral feedback (e.g., "Great! You really worked hard"). In addition, whether the feedback was very positive, positive, negative or very negative was coded. Because very positive and very negative feedback rarely occurred, we collapsed the categories into positive versus negative performance and behavioral feedback.

Based on the theoretical considerations (see sections [2.2.1](#) and [2.2.2](#)), in the present study, we investigated three characteristics of teacher feedback: (1) the *performance relatedness of feedback*, that is, the extent to

which teacher feedback was performance-related and not related to the students' behavior; (2) the *valence of performance feedback* that a child received, that is, the extent to which the performance feedback was positive and not negative; and (3) the *elaborateness of negative performance feedback*, that is, whether in the case of negative performance feedback, the teacher provided the child with further information about how to proceed. To consider differences in the duration of school lessons and in the number of interactions per child, we calculated intraindividual percentages for these three characteristics of teacher feedback.

First, the *performance relatedness of feedback* was calculated via the ratio of performance feedback to total feedback that a child received. A value of 0.5 on this measure indicates that a child received the same amount of performance feedback as of behavioral feedback. Higher values indicate that a child received more performance feedback than behavioral feedback (e.g., a value of 0.8 means that the ratio of received feedback was 80% performance and 20% behavioral feedback).

Second, the *valence of performance feedback* was calculated via the ratio of positive performance feedback to total performance feedback for each child. Values greater than 0.5 on this measure indicate that a child received more positive and therefore affirmative performance feedback than negative performance feedback.

Third, the *elaborateness of negative performance feedback* was calculated via the ratio of negative performance feedback with further elaboration to total negative performance feedback. That is, for negative performance feedback (e.g., pointing to a mistake), we determined whether the child received at least one tip or suggestion from the teacher in the same interaction. This was the case if the code "very negative/negative performance feedback" and one of the elaboration codes (i.e., direct hint/prompt/supporting question) occurred in the same interaction sequence. A higher value on this measure indicates that negative performance feedback was typically accompanied by further elaboration, while lower values point to negative performance feedback with little or no advice for improvement.

4.3.5. Background variables

Students' background characteristics can affect learning progress in school, as well as teacher expectations. To account for these influences, we considered two aspects of students' social background: socioeconomic status and parental education. The *Highest International Socio-Economic Index of Occupational Status* among the parents (HISEI; [Ganzeboom, 2010](#)) was used to represent the socioeconomic status of the family. The HISEI is an internationally established measure ([Baumert & Maaz, 2006](#); [OECD, 2003](#)) that quantifies the attributes of occupations that convert education into income ([Ganzeboom, Graaf, & Treiman, 1992](#)). Parental education was captured by a dummy-coded variable differentiating between families with at least one parent with an *Abitur* (higher education entrance qualification in Germany), coded as 1, and parents without the *Abitur*, coded as 0.

We further controlled for whether the students came from immigrant families, defined as having at least one parent born in a country other than Germany. Immigrant status was included in the analyses as a dummy variable (0 = *nonimmigrant family*, 1 = *immigrant family*). Furthermore, students' gender (0 = *male*, 1 = *female*) and students' age in years at the start of school enrollment (grand-mean centered, that is, the average student age constitutes the zero point) served as control variables when investigating students' learning progress.

4.4. Analytic strategy

4.4.1. General information

For all of the analyses we used Stata/SE data analysis software, version 15.1 ([StataCorp LLC, 1985–2017](#)). 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 percentage of missing data varied between 3 and 7% for achievement-related student variables

(achievement, ability, motivation), between 2 and 6% for teacher expectations, between 13 and 17% for student background variables, and between 1 and 16% for teacher feedback. The imputation models included not only the variables used in the analyses but also further information from teacher and parent interviews. Missing data were imputed separately for the main sample and the video subsample. All of the descriptive and regression analyses were conducted individually for 50 imputed data sets, and their parameters were subsequently pooled according to Rubin's rules (Rubin, 1987). To calculate the standardized regression coefficients, we re-ran all models with z-standardized metric variables. In all of the regression analyses, we used the cluster command in Stata (StataCorp, 2017) to consider the clustered structure of the data (students within classrooms). This procedure produces cluster-robust standard errors (which are based on the *sandwich estimator of variance* developed by Huber, 1967 and White, 1980, 1982), which account for heteroskedasticity and dependencies within nested data.

4.4.2. Separating teacher expectation inaccuracy

As a first step, we explored teacher expectation inaccuracy (Research Question 1), which we defined as teacher expectations that deviate from students' prior achievement, general cognitive abilities and motivation. To identify such inaccuracy at the beginning of the school year, we applied the residual approach proposed by Madon et al. (1997). We conducted multiple regression analyses with cluster-robust standard errors predicting teacher expectations regarding German language and mathematics performance from students' achievements (German language: phonological awareness and reading achievement; mathematics: mathematical achievement), general cognitive abilities (deductive reasoning and working memory), and motivation (enjoyment of learning and effort). The residuals resulting from these regressions reflect the variance that remained unexplained and served as our measure of teacher expectation inaccuracy. Residual scores around zero indicated accurate expectations based on the aforementioned student characteristics, positive values indicated inaccurately high expectations, and negative values indicated inaccurately low expectations.

In addition to the metric residual score, we used, in separate analyses, a series of dummy variables distinguishing between *inaccurately high expectations* (residual score more than one standard deviation greater than the mean), *inaccurately low expectations* (residual score more than one standard deviation less than the mean) and *accurate expectations* (residual score within one standard deviation of the mean). In these analyses, *accurate expectations* served as the reference category. We used a tolerance criterion of one standard deviation, as it is the changing point in normal distributions. However, as the results may differ depending on the exact criterion chosen, we additionally ran all models with 0.5 standard deviations as the cut-off for robustness checks (results presented in footnotes).

4.4.3. Analyzing teacher expectancy effects

To estimate the relationship between teacher expectation inaccuracy and teacher feedback (Research Question 2), we performed linear regression analyses with cluster-robust standard errors for each type of teacher feedback using the subsample. The models controlled for students' achievement at the beginning of the school year, general cognitive abilities, and motivation.

Subsequently, to investigate teacher expectancy effects (Research Question 3), we ran linear regression models with cluster-robust standard errors based on the main sample separately for students' reading achievement and their mathematical achievement. In addition to students' prior achievements, general cognitive abilities and motivation, in these models, we controlled for students' age, gender and family background, as these variables are known to predict students' learning progress. Since teacher expectation inaccuracy correlates with some of these student characteristics, not controlling for them would have caused the statistical relations between the teacher expectation inaccuracy and *later* student achievement to be confounded by background-related influences that existed independently from teacher

expectations (e.g., advantageous home learning environments in families with a higher socioeconomic status). In the first model, we included the metric teacher expectation inaccuracy as a predictor. As higher scores indicated either inaccurately high expectations or less inaccurately low expectations, we expected a positive regression coefficient for this variable on achievement. The second model included the dummy variables for inaccurately high and low teacher expectations, with the accurate-expectations category as the reference. This model allowed us to examine differences in the effect sizes of inaccurately high and inaccurately low teacher expectations. We expected positive regression coefficients for inaccurately high expectations and negative regression coefficients for inaccurately low expectations.

Finally, we directly investigated the assumed mediation of teacher expectancy effects through teacher feedback (Research Question 4). We explored whether teacher feedback predicted students' achievement development while controlling for teacher expectation inaccuracy, as well as whether the direct link between teacher expectation inaccuracy and student achievement significantly decreased when considering teacher feedback. Fig. 1 illustrates our overall analytic strategy.

5. Results

5.1. Teacher expectations and teacher expectation inaccuracy

Table 2 displays descriptive statistics for teachers' expectations and teachers' expectation inaccuracy in the main sample and the subsample. On average, teachers' expectations were slightly above the middle of the scale (German language: $M = 3.31$, $SD = 0.97$, mathematics: $M = 3.36$, $SD = 0.88$). That is, teachers assumed that their students performed somewhat above the classroom average overall.

Regression analyses of teacher expectations based on the main sample revealed that approximately 35% of the variation in teacher expectations was explained by students' achievements, general cognitive abilities, and motivation (German language: 33.85%, mathematics: 37.08%; models not displayed)². This result indicates, in accordance with Hypothesis 1, that approximately 65% of the variation in teacher expectations remained unexplained and was interpreted as teacher expectation inaccuracy. However, it is possible that some part of this unexplained variance was due to teacher observations of student characteristics that were not measured in this study or to measurement error. We will further debate this issue in the discussion section. According to the descriptive results displayed in Table 2, the teacher expectation inaccuracy varied substantially between students in both domains, indicating that some teacher evaluations overlapped more than others with actual student achievement, ability and motivation. The intraclass correlations showed that the major part of the variation occurred within classes (German language: $ICC = 0.08$, mathematics: $ICC = 0.05$). Descriptive statistics of the categorical teacher expectation inaccuracy variable revealed that approximately 70% of teacher expectations differed not more than one standard deviation from zero and were thus, based on this tolerance criterion, defined as accurate (German language: 66.81%, mathematics: 70.80%). At the same time, approximately 15% of teacher expectations were categorized as either inaccurately low or inaccurately high (German language: 16.04% inaccurately low, 17.14% inaccurately high; mathematics: 14.62% inaccurately low, 14.58% inaccurately high)³.

To ensure that the metric residual scores were unrelated to students'

² For further information on similar models, see Lorenz et al. (2016).

³ When the categorization was based on a tolerance criterion of 0.5 standard deviations (instead of one standard deviation), approximately 40% of teacher expectations were categorized as accurate (German language: 37.86%, mathematics: 39.99%). Correspondingly, approximately 30% were categorized as either inaccurately high (German language: 31.36%, mathematics: 30.35%) or inaccurately low (German language: 30.78%, mathematics: 29.66%).

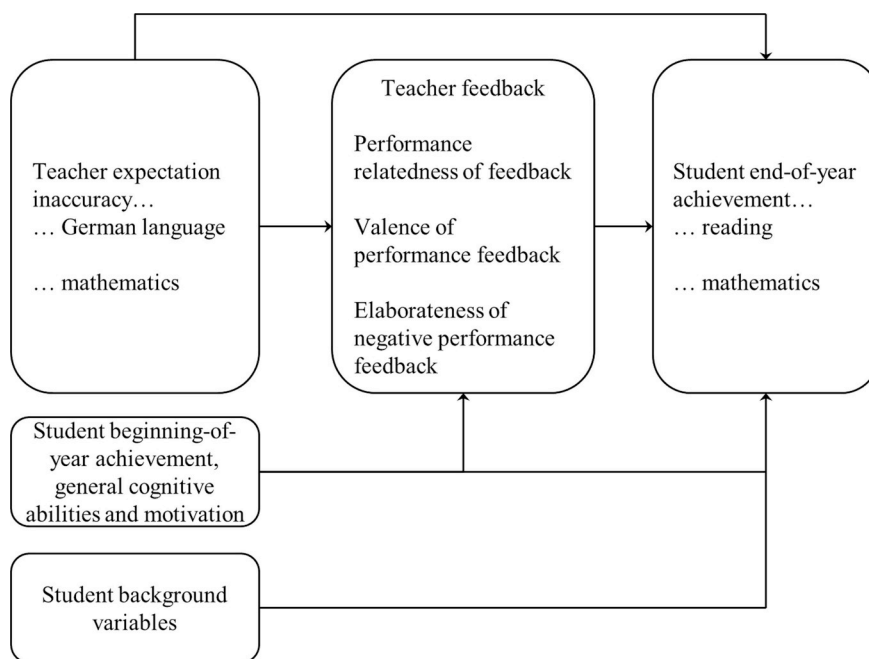


Fig. 1. Illustration of the overall analytic strategy.

initial achievements (as they should be, because initial achievement, general cognitive abilities, and motivation were controlled for in the regressions because we defined them as accurate influences on teacher expectations), we checked for bivariate correlations between the residual scores and students' beginning-of-year achievements. Except for a weak correlation in the subsample between the teacher expectation inaccuracy in the German language domain and phonological awareness ($r = 0.17, p = .002$), the correlations were all nonsignificant.

5.2. Teacher feedback

Based on the subsample, Table 2 also contains descriptive statistics of teacher feedback. During the videotaped lessons, students received, on average, individual feedback about their performance or their

behavior 23 times ($M = 23.02, SD = 18.97$). On average, 66.69% of the feedback that a student received was related to his/her performance, and 33.31% was related to his/her behavior. The ratio between performance and behavioral feedback varied substantially between students ($SD = 21.33$). Some students received only performance feedback ($max = 100.00%$), and some students received only behavioral feedback ($min = 0.00%$). Taking a closer look at the valence of performance feedback revealed that positive performance feedback was more prevalent than negative performance feedback. For a student who received at least one instance performance feedback during the videotaped lessons, the feedback was positive in 72.57% of the instances ($SD = 19.51$). Negative performance feedback accompanied by further elaboration (e.g., how to fix a mistake) occurred in 42.52% of the interactions that included at least one instance of negative performance

Table 2
Descriptive statistics of teacher expectations, teacher expectation inaccuracy and teacher feedback.

| Variables | Main sample | | | | | Subsample | | | | |
|--|-------------|------|------|-------|------|-----------|-------|-------|-------|------|
| | N | M | SD | Min | Max | n | M | SD | Min | Max |
| Teacher expectations | | | | | | | | | | |
| German language | 1026 | 3.31 | 0.97 | 1 | 5 | 354 | 3.31 | 1.00 | 1 | 5 |
| Mathematics | 1004 | 3.36 | 0.88 | 1 | 5 | 354 | 3.38 | 0.89 | 1 | 5 |
| Teacher expectation inaccuracy | | | | | | | | | | |
| German language | 1026 | 0.00 | 0.79 | -2.45 | 2.16 | 354 | 0.04 | 0.78 | -2.45 | 2.15 |
| Mathematics | 1004 | 0.00 | 0.69 | -2.47 | 2.58 | 354 | 0.04 | 0.70 | -2.47 | 2.55 |
| Categories of expectation inaccuracy | % | | | | | % | | | | |
| German language | | | | | | | | | | |
| Inaccurately low expectations | 16.04 | | | | | 15.53 | | | | |
| Accurate expectations | 66.81 | | | | | 66.11 | | | | |
| Inaccurately high expectations | 17.14 | | | | | 18.37 | | | | |
| Mathematics | | | | | | | | | | |
| Inaccurately low expectations | 14.62 | | | | | 12.92 | | | | |
| Accurate expectations | 70.80 | | | | | 69.60 | | | | |
| Inaccurately high expectations | 14.58 | | | | | 17.49 | | | | |
| Teacher feedback (FB) practice | | | | | | n | M | SD | Min | Max |
| Absolute amount of FB | | | | | | 354 | 23.02 | 18.97 | 0 | 108 |
| Performance relatedness of FB | | | | | | 351 | 66.69 | 21.33 | 0 | 100 |
| Valence of performance FB | | | | | | 348 | 72.57 | 19.51 | 0 | 100 |
| Elaborateness of negative performance FB | | | | | | 297 | 42.52 | 34.34 | 0 | 100 |

Table 3
Regression of three characteristics of teacher feedback (FB) on teacher expectation inaccuracy.

| Variable | Performance relatedness of FB | | | | Valence of performance FB | | | | Elaborateness of negative performance FB | | | |
|-------------------------------------|-------------------------------|------------------|-----------------|---------|---------------------------|---------|-----------------|------------------|--|-------------------|-----------------|---------|
| | Model 1 | | Model 2 | | Model 3 | | Model 4 | | Model 5 | | Model 6 | |
| | <i>b</i> (SE) | β | <i>b</i> (SE) | β | <i>b</i> (SE) | β | <i>b</i> (SE) | β | <i>b</i> (SE) | β | <i>b</i> (SE) | β |
| Teacher expectation inaccuracy | | | | | | | | | | | | |
| German language | 3.39 (1.44) | .12* | | | 2.58 (1.12) | .10* | | | -5.03 (2.43) | -.11 ⁺ | | |
| Mathematics | | | 3.95 (1.84) | .13* | | | 2.84 (1.11) | .10* | | | -3.58 (2.81) | -.07 |
| Student achievement t0 | | | | | | | | | | | | |
| Phonological awareness | 0.09 (0.23) | .02 | | | 0.40 (0.31) | .09 | | | -0.27 (0.67) | -.04 | | |
| Reading | 0.02 (0.06) | .03 | | | 0.13 (0.05) | .15* | | | 0.02 (0.12) | .01 | | |
| Mathematics | | | -0.06 (0.20) | -.02 | | | 0.32 (0.17) | .14 ⁺ | | | -0.43 (0.27) | -.11 |
| Student general cognitive abilities | | | | | | | | | | | | |
| Deductive reasoning | 0.78 (0.45) | .11 | 0.92 (0.55) | .13 | 0.21 (0.38) | .03 | 0.30 (0.39) | .05 | -1.05 (0.79) | -.09 | -0.62 (0.87) | -.05 |
| Working memory | 0.79 (0.67) | .06 | 0.85 (0.64) | .07 | 0.13 (0.82) | .01 | -0.07 (0.82) | -.01 | -3.09 (1.08) | -.15* | -2.71 (1.01) | -.13* |
| Student motivation | | | | | | | | | | | | |
| Enjoyment of learning | 8.38 (4.31) | .15 ⁺ | 8.77 (4.09) | .15* | 0.38 (3.40) | .01 | 0.97 (3.41) | .02 | -8.16 (6.59) | -.09 | -8.25 (6.54) | -.09 |
| Effort | -2.64 (6.58) | -.03 | -2.19 (6.56) | -.03 | 4.20 (4.09) | .06 | 4.63 (4.05) | .07 | 7.75 (9.21) | .06 | 8.98 (8.87) | .07 |
| Intercept | 49.08*** (8.05) | | 49.88*** (7.66) | | 55.05*** (6.05) | | 52.84*** (5.41) | | 61.20*** (9.33) | | 63.45*** (8.98) | |
| R-squared | 6.26% | | 6.23% | | 7.93% | | 5.25% | | 5.66% | | 5.48% | |
| <i>n</i> | 351 | | 351 | | 348 | | 348 | | 297 | | 297 | |
| Cluster | 19 | | 19 | | 19 | | 19 | | 19 | | 19 | |

Note. t0 = beginning of first grade. ⁺ $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

feedback ($SD = 34.34$). Table 2 also reveals that the sample sizes for the three characteristics of teacher feedback varied. This variation is due to the fact that some students received no feedback ($n = 3$), no performance feedback ($n = 3$) or no negative performance feedback ($n = 51$) during the video-recorded school lessons. Thus, for these students, we could not calculate intraindividual percentages (see section 4.3.4 for more information) and had to exclude them from the respective analyses.

5.2.1. Teacher expectation inaccuracy and teacher feedback

Based on the subsample, we investigated the relation between teacher expectation inaccuracy and teacher feedback. The results of the multivariate regressions with cluster-robust standard errors displayed in Table 3 indicate positive relationships between teacher expectation inaccuracy and two of the three characteristics of teacher feedback, even after students' achievement, general cognitive abilities, and motivation were controlled. In accordance with Hypotheses 2a and 2b, higher values on the variable of teacher expectation inaccuracy (that is, inaccurately higher expectations or less inaccurately low expectations) were associated with a higher intraindividual ratio of performance feedback than of behavioral feedback (German language: $\beta = 0.12$, $p = .031$; mathematics: $\beta = 0.13$, $p = .047$) and with more positive than negative performance feedback (German language: $\beta = 0.10$, $p = .035$; mathematics: $\beta = 0.10$, $p = .021$). That is, if the value of the teacher expectation inaccuracy increased by 1 point, approximately 4% more of the feedback a student received was performance feedback and not behavioral feedback; similarly, the percentage of positive performance feedback compared to negative performance feedback increased by approximately 3% when the teacher expectation inaccuracy increased by 1 point. In other words, regardless of students' actual achievements, teachers tended to give more performance feedback (in comparison to behavioral feedback) and more positive than negative performance feedback to students whom they inaccurately expected to show higher achievement. Conversely, equally achieving lower-expectancy students received less performance feedback and more behavioral feedback from their teachers, and the performance feedback was less likely to be positive. Contradicting Hypothesis 2c, teachers did not provide more elaboration in the case of negative performance feedback for their inaccurately higher-expectancy students. On the contrary, the results tend to point in the opposite direction, indicating more elaboration for inaccurately lower-expectancy students. However, this weak association is significant on the 10%-level only in the German

language domain (German language: $\beta = -.11$, $p = .055$; mathematics: $\beta = -.07$, $p = .222$).

5.3. Teacher expectancy effects on student learning

Based on the main sample, we further examined the direct link between teacher expectation inaccuracy and student learning (see Table 4). In the German language domain, the metric residual expectancy score significantly predicted reading achievement at the end of the school year (model 1, $\beta = 0.21$, $p < .001$). Consistent with Hypothesis 3, this result indicates that higher scores on teacher expectation inaccuracy (that is, inaccurately higher expectations or less inaccurately low expectations) in the German language domain were associated with higher reading skills at the end of first grade, even after students' prior achievements, general cognitive abilities, motivation, gender, age, and family backgrounds were considered. The variation in the end-of-school-year reading achievement that was additionally explained by teacher expectation inaccuracy amounted to 4.07% (model without teacher expectation inaccuracy not displayed: $R^2 = 25.60\%$; model 1: $R^2 = 29.67\%$). We further investigated the regression coefficients of inaccurately high expectations and the coefficients of inaccurately low expectations, while accurate expectations served as the reference category. Model 2 confirmed that, compared to accurate expectations, inaccurately high expectations were related to higher end-of-year reading achievement ($\beta = 0.25$, $p = .002$), and inaccurately low expectations were related to lower ($\beta = -.42$, $p < .001$) end-of-year reading achievement.

In mathematics, the association between teacher expectation inaccuracy and student achievement was somewhat less pronounced. Teacher expectation inaccuracy significantly predicted gains in end-of-year mathematical achievement (model 3, $\beta = 0.09$, $p = .001$) when controlling for students' prior achievements, general cognitive abilities, motivation, gender, age, and family background (0.62% of the variance was additionally explained by the teacher expectation inaccuracy; model without teacher expectation inaccuracy not displayed: $R^2 = 57.13\%$; model 3: $R^2 = 57.75\%$). Moreover, comparing inaccurately high and inaccurately low expectations to accurate expectations revealed only small differences in students' achievements (model 4). Consistent with Hypothesis 3a, students who were exposed to inaccurately high expectations gained somewhat more mathematical skills during first grade ($\beta = 0.14$, $p = .014$). However, in contrast to Hypothesis 3b, students for whom their teachers had inaccurately low

Table 4
Teacher expectation inaccuracy as a predictor of student achievement.

| Variable | Reading | | | | Mathematics | | | |
|-------------------------------------|------------------------|---------|------------------------|---------|------------------------|---------|------------------------|---------|
| | Model 1 | | Model 2 | | Model 3 | | Model 4 | |
| | <i>b</i> (<i>SE</i>) | β | <i>b</i> (<i>SE</i>) | β | <i>b</i> (<i>SE</i>) | β | <i>b</i> (<i>SE</i>) | β |
| Teacher expectation inaccuracy | | | | | | | | |
| German language | | | | | | | | |
| Metric inaccuracy | 9.91 (1.79) | .21*** | | | | | | |
| Inaccurately high expectations | | | 9.35 (2.81) | .25** | | | | |
| Inaccurately low expectations | | | -15.93 (3.69) | -.42*** | | | | |
| Mathematics | | | | | | | | |
| Metric inaccuracy | | | | | 0.89 (0.25) | .09*** | | |
| Inaccurately high expectations | | | | | | | 0.98 (0.39) | .14* |
| Inaccurately low expectations | | | | | | | -0.71 (0.53) | -.10 |
| Student achievement t0 | | | | | | | | |
| Phonological awareness | 1.93 (0.28) | .24*** | 1.87 (0.28) | .23*** | | | | |
| Reading | 0.05 (0.04) | .03 | 0.04 (0.04) | .03 | | | | |
| Mathematics | | | | | 0.54 (0.03) | .62*** | 0.54 (0.03) | .62*** |
| Student general cognitive abilities | | | | | | | | |
| Deductive reasoning | 2.05 (0.41) | .16*** | 2.04 (0.41) | .16*** | 0.28 (0.06) | .11*** | 0.28 (0.06) | .11*** |
| Working memory | 1.26 (0.55) | .06* | 1.24 (0.57) | .06* | 0.06 (0.09) | .02 | 0.07 (0.09) | .02 |
| Student Motivation | | | | | | | | |
| Enjoyment of learning | 3.28 (4.30) | .03 | 3.26 (4.47) | .03 | 0.69 (0.60) | .04 | 0.71 (0.61) | .04 |
| Effort | 23.75 (5.99) | .18*** | 23.56 (6.03) | .18*** | 0.96 (0.90) | .04 | 0.86 (0.91) | .03 |
| Student background | | | | | | | | |
| Student age | -0.54 (3.66) | -.00 | -0.53 (3.67) | -.00 | -0.55 (0.51) | -.03 | -0.52 (0.51) | -.02 |
| Female gender | 0.86 (1.90) | .02 | 1.40 (1.90) | .04 | -1.79 (0.32) | -.25*** | -1.79 (0.32) | -.26*** |
| HISEI | 0.16 (0.07) | .08* | 0.17 (0.07) | .09* | 0.01 (0.01) | .04 | 0.01 (0.01) | .04 |
| Abitur | 0.49 (2.68) | .01 | 1.00 (2.69) | .03 | 0.36 (0.36) | .05 | 0.43 (0.36) | .06 |
| Immigrant status | -0.21 (2.78) | -.01 | -0.19 (2.70) | -.00 | -0.17 (0.36) | -.02 | -0.14 (0.36) | -.02 |
| Intercept | 9.16 (9.01) | | 10.65 (9.08) | | 19.46*** (1.23) | | 19.43*** (1.22) | |
| R-squared | 29.67% | | 29.27% | | 57.75% | | 57.43% | |
| N | 1026 | | 1026 | | 1004 | | 1004 | |
| Cluster | 64 | | 64 | | 63 | | 63 | |

Note. t0 = beginning of first grade. + $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

expectations did not differ from their accurately estimated classmates ($\beta = -.10, p = .188$).⁴

5.4. Mediating role of teacher feedback

To investigate the mediating role of teacher feedback in the expectancy effects identified above, we first examined whether teacher feedback predicted student achievement in reading and mathematics at the end of the school year after controlling for teacher expectation inaccuracy, as well as students' prior achievements, general cognitive abilities, motivation, gender, age, and family background (see Tables 5 and 6). The multivariate regression analyses with cluster-robust standard errors were based on the subsample. The analyses revealed a positive association between the valence of performance feedback and mathematical achievement at the end of the school year ($\beta = 0.08, SE = 0.03, p = .020, 95\% CI [0.01, 0.14]$); that is, the more positive and the less negative performance feedback a student received, the more mathematical skills the student gained during first grade. In reading, the valence of performance feedback did not predict students' learning ($\beta = 0.04, SE = 0.04, p = .382, 95\% CI [-0.05, 0.13]$). This result was also observed for the other two indicators of teacher feedback: no significant regression coefficients were found for the

performance relatedness of feedback (reading: $\beta = 0.07, SE = 0.07, p = .307, 95\% CI [-0.08, 0.22]$; mathematics: $\beta = 0.02, SE = 0.04, p = .547, 95\% CI [-0.06, 0.11]$) and the elaborateness of negative performance feedback (reading: $\beta = -.02, SE = 0.06, p = .730, 95\% CI [-0.15, 0.11]$; mathematics: $\beta = 0.02, SE = 0.04, p = .607, 95\% CI [-0.07, 0.11]$) for reading or mathematical achievement.

Further, we investigated the potential mediation of the teacher expectancy effect in the German language (Table 5) and the mathematical domain (Table 6) by the three characteristics of teacher feedback. We compared the coefficient of teacher expectation inaccuracy in model 0, which shows the expectancy effect in the subsample, to the coefficients of teacher expectation inaccuracy in models 1, 2 and 3 in the same table, each of which additionally controlled for teacher feedback. For all three teacher feedback variables, the results did not indicate a substantial mediation of teacher expectancy effects, which contradicted Hypothesis 4 (for all changes in the coefficient of teacher expectation inaccuracy, $p > .10$). For example, the direct effect of inaccurate teacher expectations on mathematical end-of-year achievement decreased marginally when additionally controlling for the valence of performance feedback, from $b = 0.97, \beta = 0.10, SE = 0.04, p = .030, 95\% CI [0.01, 0.18]$ to $b = 0.93, \beta = 0.09, SE = 0.04, p = .041, 95\% CI [0.00, 0.18]$. The reduction was not statistically significant.

Based on considerations by Zhao, Lynch, and Chen (2010), we additionally examined the indirect effects of teacher expectation inaccuracy on student achievement via each of the three teacher feedback variables in order to examine mediation. We calculated the indirect effects with structural equation models using a full information maximum likelihood procedure. Since the standard errors of indirect effects are non-normally distributed (Jose, 2013), we applied a Monte Carlo approach using the Stata package "medsem" (Mehmetoglu, 2018) in

⁴ When the categorization of the teacher expectation inaccuracy was based on a tolerance criterion of 0.5 standard deviations (instead of one standard deviation), the results point in the same direction, although the effect sizes were somewhat lower (German language: inaccurately high expectation: $\beta = 0.17, p = .039$, inaccurately low expectation: $\beta = -.32, p < .001$; mathematics: inaccurately high expectation: $\beta = 0.14, p = .005$, inaccurately low expectation: $\beta = -.08, p = .188$).

Table 5
Mediation of the relationship between teacher expectation inaccuracy and students' reading achievement.

| Variable | Model 0 | | Model 1 | | Model 2 | | Model 3 | |
|--|---------------|------------------|---------------|------------------|---------------|---------|---------------|------------------|
| | <i>b</i> (SE) | β | <i>b</i> (SE) | β | <i>b</i> (SE) | β | <i>b</i> (SE) | β |
| Teacher expectation inaccuracy | | | | | | | | |
| German language | 8.50 (3.25) | .19* | 8.37 (3.32) | .18* | 8.63 (3.39) | .19* | 9.20 (3.30) | .20* |
| Teacher feedback (FB) | | | | | | | | |
| Performance relatedness of FB | | | 0.13 (0.12) | .07 | | | | |
| Valence of performance FB | | | | | 0.07 (0.08) | .04 | | |
| Elaborateness of negative performance FB | | | | | | | -0.02 (0.06) | -.02 |
| Student achievement t0 | | | | | | | | |
| Phonological awareness | 2.36 (0.53) | .29*** | 2.41 (0.53) | .30*** | 2.38 (0.52) | .30*** | 2.32 (0.48) | .29*** |
| Reading | -0.03 (0.07) | -.02 | -0.04 (0.07) | -.02 | -0.04 (0.07) | -.03 | -0.01 (0.08) | -.01 |
| Student general cognitive ability | | | | | | | | |
| Deductive reasoning | 1.10 (0.56) | .09 ⁺ | 1.07 (0.55) | .09 ⁺ | 1.23 (0.57) | .10* | 1.28 (0.67) | .11 ⁺ |
| Working memory | 2.58 (0.92) | .12* | 2.48 (0.91) | .11* | 2.51 (0.92) | .12* | 2.11 (1.02) | .10 ⁺ |
| Student motivation | | | | | | | | |
| Enjoyment of learning | -6.74 (6.94) | -.07 | -6.27 (7.02) | -.07 | -5.50 (6.87) | -.06 | -4.42 (8.56) | -.05 |
| Effort | 30.58 (10.04) | .24** | 27.76 (10.09) | .22* | 27.05 (9.96) | .21* | 29.24 (10.80) | .23* |
| Student background | | | | | | | | |
| Student age | 4.69 (7.55) | .05 | 4.39 (7.86) | .04 | 4.70 (7.98) | .05 | 6.52 (8.18) | .06 |
| Female gender | 3.62 (3.17) | .10 | 2.21 (3.56) | .06 | 3.54 (3.46) | .10 | 3.02 (3.39) | .08 |
| HISEI | 0.15 (0.12) | .08 | 0.17 (0.11) | .10 | 0.17 (0.11) | .09 | 0.13 (0.12) | .07 |
| Abitur | -3.18 (4.52) | -.09 | -3.64 (4.38) | -.10 | -2.85 (4.71) | -.08 | -1.19 (5.09) | -.03 |
| Immigrant status | 2.06 (3.85) | .06 | 2.48 (3.98) | .07 | 2.40 (4.01) | .07 | 2.63 (3.96) | .07 |
| Intercept | 13.02 (14.54) | | 8.30 (13.54) | | 10.15 (15.31) | | 12.15 (14.26) | |
| R-squared | 31.17% | | 31.43% | | 31.42% | | 31.76% | |
| <i>n</i> | 354 | | 351 | | 348 | | 297 | |
| Cluster | 19 | | 19 | | 19 | | 19 | |

Note. t0 = beginning of first grade. ⁺*p* < .10. * *p* < .05. ** *p* < .01. *** *p* < .001.

Table 6
Mediation of the relationship between teacher expectation inaccuracy and students' mathematical achievement.

| Variable | Model 0 | | Model 1 | | Model 2 | | Model 3 | |
|--|-----------------|------------------|-----------------|------------------|-----------------|---------|-----------------|-------------------|
| | <i>b</i> (SE) | β | <i>b</i> (SE) | β | <i>b</i> (SE) | β | <i>b</i> (SE) | β |
| Teacher expectation inaccuracy | | | | | | | | |
| Mathematics | 0.97 (0.40) | .10* | 0.97 (0.43) | .10* | 0.93 (0.42) | .09* | 0.93 (0.43) | .09* |
| Teacher feedback (FB) | | | | | | | | |
| Performance relatedness of FB | | | 0.01 (0.01) | .02 | | | | |
| Valence of performance FB | | | | | 0.03 (0.01) | .08* | | |
| Elaborateness of negative performance FB | | | | | | | 0.00 (0.01) | .02 |
| Student achievement t0 | | | | | | | | |
| Mathematics | 0.54 (0.05) | .66*** | 0.54 (0.05) | .66*** | 0.53 (0.05) | .64*** | 0.55 (0.05) | .67*** |
| Student general cognitive abilities | | | | | | | | |
| Deductive reasoning | 0.20 (0.09) | .08* | 0.20 (0.09) | .09* | 0.21 (0.09) | .09* | 0.25 (0.10) | .11* |
| Working memory | 0.10 (0.17) | .02 | 0.09 (0.17) | .02 | 0.09 (0.17) | .02 | 0.14 (0.18) | .03 |
| Student motivation | | | | | | | | |
| Enjoyment of learning | 0.07 (0.87) | .00 | 0.09 (0.88) | .00 | 0.17 (0.87) | .01 | -0.13 (0.93) | -.01 |
| Effort | 1.22 (1.61) | .05 | 1.14 (1.67) | .05 | 1.02 (1.66) | .04 | 1.22 (1.66) | .05 |
| Student background | | | | | | | | |
| Student age | -0.53 (0.60) | -.03 | -0.56 (0.61) | -.03 | -0.57 (0.63) | -.03 | -0.20 (0.69) | -.01 |
| Female gender | -1.06 (0.49) | -.15* | -1.17 (0.51) | -.17* | -1.25 (0.51) | -.18* | -0.96 (0.49) | -.14 ⁺ |
| HISEI | 0.03 (0.02) | .10 ⁺ | 0.03 (0.02) | .09 ⁺ | 0.03 (0.02) | .09 | 0.03 (0.02) | .09 ⁺ |
| Abitur | -0.64 (0.53) | -.09 | -0.68 (0.53) | -.10 | -0.62 (0.54) | -.09 | -0.63 (0.57) | -.09 |
| Immigrant status | 0.08 (0.64) | .01 | 0.09 (0.65) | .01 | -0.01 (0.64) | -.00 | 0.19 (0.69) | .03 |
| Intercept | 19.48*** (2.31) | | 19.29*** (2.33) | | 18.41*** (2.45) | | 18.73*** (2.50) | |
| R-squared | 59.69% | | 59.40% | | 59.76% | | 59.93% | |
| <i>n</i> | 354 | | 351 | | 348 | | 297 | |
| Cluster | 19 | | 19 | | 19 | | 19 | |

Note. t0 = beginning of first grade. ⁺*p* < .10. * *p* < .05. ** *p* < .01. *** *p* < .001.

order to test for the significance of the indirect effects. The results (models not displayed) confirmed the insignificance of the indirect effects and therefore indicate direct-only nonmediation for all three indicators of teacher feedback examined in this study.

6. Discussion

Contributing to the body of research on the role of teachers' expectations in students' academic achievements, this study focused on the mediating mechanisms and empirically traced all three steps of the

assumed self-fulfilling prophecy process. Our first goal was to clarify whether associations between teacher expectations and later student achievement occurred due to differences in teacher behavior. The second aim was to resolve open questions surrounding the role of teacher feedback as a mediator of teacher expectancy effects. In this regard, we examined the associations between teacher expectations, teacher feedback practice and student learning.

In summary, the results showed that teacher expectations were inaccurate to some extent. Further, this inaccuracy in teacher expectations significantly predicted students' end-of-year achievement, even when prior achievement, general cognitive abilities, motivation, and student background characteristics were controlled. Specifically, inaccurately high expectations were related to greater achievement gains in reading and mathematics, whereas inaccurately low expectations were related to lower achievement gains in reading only. In addition, teacher feedback varied with teacher expectations but did not substantially mediate teacher expectancy effects.

The results are in line with existing findings indicating that teacher expectations are partly inaccurate (Research Question 1). That is, in accordance with Hypothesis 1, teacher expectations differ from actual student achievement, general cognitive abilities, and motivation. The shared variance between student characteristics and teacher expectations in the German language domain and in mathematics amounted to approximately 35%, corresponding to estimates reported in the available meta-analyses (Hoge & Coladarsi, 1989; Südkamp et al., 2012). When using dichotomous inaccuracy categories, we defined teacher expectations to be accurate in a range of one standard deviation below and above the mean. This approach resulted in 65–70% of teacher expectations being defined as accurate. However, the exact proportion depended on the tolerance range chosen. With a tolerance range of 0.5 standard deviations, for example, only 40% of the teacher expectations were categorized as accurate. The exact proportions of the categories should thus be interpreted with caution. As the robustness checks of all analyses with different tolerance ranges of teacher expectation accuracy revealed, the results and the conclusions did not differ substantially depending on the definition of this criterion.

Teacher expectation inaccuracy was also significantly associated with two dimensions of teacher feedback (Research Question 2). In line with Hypotheses 2a and 2b, compared to similar-achieving lower-expectancy classmates, higher-expectancy students received more performance feedback than behavioral feedback and somewhat more positive performance feedback than negative performance feedback. This finding generally supports the assumption that teachers communicate their expectations through different feedback practices. The frequency of further elaboration provided by teachers in the case of negative performance feedback, however, did not increase with inaccurately higher expectations, which contradicted Hypothesis 2c.

The results further indicate that teacher expectations can result in a self-fulfilling prophecy (Research Question 3). Associations between initial teacher expectation inaccuracy and later student achievements were found for the German language and mathematical domains, but the regression coefficient was larger in the language domain ($\beta = 0.21$ compared to $\beta = 0.09$). Correspondingly, whereas we found positive and negative teacher expectancy effects in the language domain, we observed only positive expectancy effects in mathematics. Therefore, the results support Hypothesis 3a for both domains but support Hypothesis 3b for the German language domain only. There could be various reasons for the differences between domains. One possible explanation might be that the linguistic domain provides more room for interpretation in the evaluation of achievement. The scope for interpretation is smaller in the case of mathematics, in which most tasks have objectively correct and incorrect responses. Another reason might be related to the different levels of preknowledge that students bring with them when entering first grade. Whereas many children can successfully count to ten or twenty and have mastered the first rules of arithmetic (cf. Deutscher & Selter, 2013), only a few already can read

when they enter school (cf. Juska-Bacher, 2013). As a result, the growth in reading achievement is much steeper than that in mathematics, and teachers (and their expectations) might therefore have a stronger impact on reading development.

The sizes of teacher expectancy effects ($\beta = 0.21$ in the German language domain and $\beta = 0.09$ in mathematics) correspond closely to the findings of earlier studies (effect sizes of 0.10–0.20; see, e.g., Jussim et al., 2009). According to the effect size guidelines for research on individual differences based on 708 meta-analyses, the effects were small to medium (Gignac & Szodorai, 2016). However, it seems necessary to consider the short time period of one school year covered in the empirical study. Teacher expectancy effects might accumulate and become stronger over longer periods (Rubie-Davies et al., 2014). Furthermore, since teacher expectations have been found to be systematically biased toward different groups of students (based on student socioeconomic background, country of origin, gender, or disability status; e.g., Lorenz et al., 2016; Shifrer, 2016), the observed expectancy effects might contribute to educational inequalities.

The mediation analyses yielded no signs of a strong mediation of teacher expectancy effects by teacher feedback (Research Question 4) and thus contradicted Hypothesis 4. Following the classification of mediation and nonmediation by Zhao et al. (2010), the results indicate direct-only nonmediation. Apart from a positive association between positive performance feedback and mathematical achievement gains, no significant correlations between teacher behavior and student learning were found. Nevertheless, the results support the association of teacher expectations with teacher feedback. One reason why the analyzed characteristics of teacher feedback did not predict student achievement gains could be the difficulty of adequately measuring high-quality feedback. Another reason for the weak support for a mediation of the effects of teacher expectations by teacher feedback might be that feedback is one of numerous channels through which teachers communicate their expectations.

6.1. Limitations and future research

Although the design of the current study allowed us to address several shortcomings of earlier research, the study also has limitations that should be addressed in future research. One limitation concerns the sample. Since teachers and students participated in the study voluntarily, the sample might be selective in terms of teacher engagement and students' socioeconomic background. A comparison of the average HISEI values in the sample ($M = 52.51$) to the population average in North Rhine-Westphalia ($M = 48.10$; Richter, Kuhl, & Pant, 2012) revealed a positive bias in the student sample. Regarding teachers, one would expect highly engaged teachers to participate more often in such a study and to be particularly eager to evaluate students accurately. As a result, the extent of inaccuracy in teacher expectations and its association with student learning might be somewhat underestimated in our study.

In addition, as the current study examined data from teachers and students in German primary schools, the conclusions are restricted to this context, more precisely, to first grade classrooms in North Rhine-Westphalia. Replication studies based on data from other countries, different grade levels (e.g., after a school transition) or specific groups of students (e.g., with different levels of self-perceptions of their ability) are important next steps to better understand the processes underlying teacher expectancy effects and ensure external validity of the current findings.

Furthermore, the subsample including information on teacher behavior was rather small. Only 19 teachers (with 354 students) agreed to be videotaped during lessons. In addition, based on the intraindividual measures of teacher feedback, we had to exclude some students from the analyses because they did not receive any feedback, performance feedback or negative performance feedback during the videotaped lessons. Therefore, the statistical power of the analyses was limited, and

we could use only a restricted number of explanatory variables in the models because of the limited degrees of freedom. Additionally, we could observe only extracts of teacher-student interactions. This limitation applies to the types of interactions, as well as to their duration. Specifically, we could examine only three characteristics of teacher feedback in four school lessons per class and were not able to consider the quality and appropriateness of students' preceding answers and behaviors. The main reason for this limitation is that video ratings are complex and expensive. However, in light of these restrictions, it is all the more remarkable that the analyses revealed statistically significant associations among teacher expectations, teacher behavior and, partly, student learning. Nevertheless, further studies should focus on the mediating mechanisms with larger sample sizes and additional dimensions of teacher behavior, as well as finer-grained measures of high-quality teacher feedback. Additional dimensions of teacher behavior may include aspects covered by the four-factor model (Rosenthal, 1973), teacher behaviors that are discussed in the context of enhancing student motivation (e.g., TARGET framework; Jussim et al., 2009) as well as other channels for the mediation of teacher expectations such as non-verbal communication (Babad, 2009). Finer-grained ratings of high-quality feedback should, for example, provide more detailed information about the three aspects of effective feedback, that is, whether the feedback contains information about (1) the learning goal, (2) how well a student is doing, and (3) the next steps to be undertaken to reach the learning goal (Hattie & Timperley, 2007). In addition, ratings should include information about the quality of students' preceding answers and behaviors.

A further limitation concerns the residual approach that was applied to identify teacher expectation inaccuracy. The unexplained variance in teacher expectations might not necessarily reflect inaccuracy but could at least partly be the result of measurement error or differences in unobserved student characteristics. First, with regard to measurement error, the risk of overestimating inaccuracy cannot be fully ruled out in field studies because not even the best instruments to assess learning-related student characteristics will represent students' true values of achievement, general cognitive abilities or motivation without error. However, as long as the potential misestimations varied randomly around the true values, the problem of measurement error should not have biased the examined relationships between teacher expectations, teacher feedback and student achievement systematically. Second, the unexplained variance in teacher expectations could partly be the result of unobserved student characteristics. If, for example, teachers had observed differences in student achievement and skills not captured by the various instruments used in this study, the residuals would not represent teacher expectation inaccuracy; instead, they could reflect accurate influences of unobserved student characteristics (cf. Lorenz, 2018). However, the various learning-related student characteristics (including general cognitive abilities and motivation) and the way in which the data were collected in our study should have minimized this problem. Since teacher-student contact prior to the first survey wave was minimal, it is fair to assume not only that teacher expectations had not yet affected initial student achievement but also that teacher expectations had not yet been determined by comprehensive evaluations of unmeasured individual student characteristics. This assumption is supported by a study that showed that ethnic bias in teacher expectations varied with teachers' stereotypes (Lorenz, 2019). Thus, we can assume that a substantial portion of the residual scores for teacher expectations emerged due to teachers' application of schemata and not due to the processing of individual student information.

6.2. Conclusion

Our study provides evidence for the relevance of teacher expectancy

effects for students' learning during the first grade in two different school domains and, in part, for the communication of teacher expectations through teachers' feedback practices. Although the examined characteristics of teacher feedback did not mediate teacher expectancy effects, the effectiveness of high-quality feedback in supporting student learning remains important (e.g., Hattie, 2009). High-quality feedback is closely related to learning goals and informs students about their progress as well as about how to proceed to reach their goals (Hattie & Timperley, 2007). To be effective, such feedback requires that teachers and students regularly set appropriate goals together. Such goals should be specific, clear and challenging but achievable; they should be reviewed and updated regularly with each student (Clarke, Timperley, & Hattie, 2003). Based on our results, teacher trainings that focus on formative high-quality feedback should enable teachers to provide supporting feedback to all their students, independent of what achievement they perceive or expect from these students.

Furthermore, our findings support the idea that students learn best if teacher demands are slightly higher than students' actual skills. This finding implies that high expectations might be beneficial for all students. Therefore, it seems worthwhile to inform teachers not only about expectancy effects in general but also about the positive effects of high expectations in particular. In addition, teachers may be encouraged to form high expectations for all their students. This is particularly important, as teacher expectancy effects reflect mostly unconscious processes. Hence, the prevention of such effects might not be entirely realistic. A central challenge in the implementation of high expectations for all students, however, may be to reduce biasing influences associated with students' gender, family backgrounds and disabilities, as current research indicates group-specific biases in teacher expectations (e.g., Hurwitz et al., 2007; Lorenz et al., 2016). This approach involves knowledge about stereotypes and strategies to reduce their influence when forming expectations.

An intervention study by Rubie-Davies, Peterson, Sibley, and Rosenthal (2015) showed that a focus on high overall expectations for all students might be a promising route. In four workshops spread over two months, the authors trained the participating teachers in the practices of teachers who had high expectations for all their students (e.g., providing goal-related feedback; Rubie-Davies et al., 2007). These workshops were supplemented by periods in which the teachers implemented the learned practices in their classrooms, self-analyses of their video-taped classroom practices and follow-up meetings with the researchers as project partners. After the intervention program, the students who were taught by teachers from the intervention group had gained more competencies in mathematics over the year than the students from the control group. This and similar training programs are valuable and promising attempts to benefit from the positive effects of high expectations and may support teachers in providing high-quality feedback to all of their students.

Funding

The authors acknowledge financial support from the Federal Ministry of Education and Research of Germany (project number 01JC1117). The responsibility for the content of this publication lies with the authors.

Author note

We would like to thank Petra Stanat for her valuable comments on an earlier version of this manuscript. Her remarks helped very much to further improve the text.

Appendix A. Descriptive statistics of all variables in the main sample and the subsample

| Variable | Main sample (N = 1026) | | | | Subsample (n = 354) | | | |
|--|------------------------|-------|-------|-------|---------------------|-------|-------|-------|
| | M | SD | Min | Max | M | SD | Min | Max |
| Student achievement t0 | | | | | | | | |
| Phonological awareness | 13.25 | 4.64 | 1 | 26 | 13.40 | 4.45 | 2 | 26 |
| Reading ^a | 23.22 | 26.43 | 0 | 156 | 20.15 | 23.46 | 0 | 147 |
| Mathematics | 27.00 | 8.06 | 0 | 53 | 26.62 | 8.43 | 0 | 53 |
| Student achievement t2 | | | | | | | | |
| Phonological awareness | 20.82 | 4.61 | 3 | 26 | 20.77 | 4.55 | 7 | 26 |
| Reading | 107.47 | 37.65 | 3 | 158 | 109.16 | 35.81 | 7 | 158 |
| Mathematics | 38.61 | 7.03 | 17 | 53 | 38.71 | 6.94 | 19 | 53 |
| Student general cognitive abilities | | | | | | | | |
| Deductive reasoning | 6.61 | 2.90 | 0 | 12 | 6.50 | 2.96 | 0 | 12 |
| Working memory | 3.11 | 1.70 | 0 | 8 | 2.88 | 1.66 | 0 | 7 |
| Student motivation | | | | | | | | |
| Enjoyment of learning | 1.53 | 0.38 | 0 | 2 | 1.54 | 0.37 | 0 | 2 |
| Effort | 1.69 | 0.28 | 0.31 | 2 | 1.69 | 0.28 | 0.62 | 2 |
| Student background | | | | | | | | |
| Age | 6.45 | 0.33 | 5.58 | 8.08 | 6.48 | 0.35 | 5.67 | 8 |
| HISEI | 52.51 | 19.44 | 14.21 | 88.96 | 52.03 | 19.69 | 14.21 | 88.96 |
| % | | | | | % | | | |
| Gender: Female | 48.11 | | | | 47.18 | | | |
| Parental education: Abitur or higher | 49.41 | | | | 49.86 | | | |
| Immigrant status | 35.55 | | | | 38.40 | | | |

Note. Descriptive statistics are based on the imputed data (pooled according to Rubin's rules; Rubin, 1987). For descriptive statistics on teacher expectations, teacher expectation inaccuracy and teacher feedback, see Table 2 t0 = beginning of first grade. t2 = end of first grade.

^a Because there are usually some students who have already learned how to read before they entered school, whereas most students have not, the reading variable naturally shows a very large standard deviation at the beginning of first grade. Robustness checks excluding this variable from the regression models, displayed in Tables 4–6, confirmed that the results are reliable.

Appendix B. Excerpt from the teacher questionnaire measuring teacher expectations (cf. Gentrup, Rjosk, Stanat, & Lorenz, 2018)

| Please evaluate how the child is likely to develop with regard to the following skills during the upcoming school year by comparing the child to the class average. | | | | | | |
|---|---|---|---------------------------------------|---------------------------------------|---------------------------------------|---------------------------------------|
| | | The child will perform in this area ... | | | | |
| | | much worse than... | a little worse than ... | just as well as ... | slightly better than ... | much better than... |
| | | ... the class average. | | | | |
| a) | linguistic skills (e.g., vocabulary and sentence structure) | <input type="checkbox"/> ₁ | <input type="checkbox"/> ₂ | <input type="checkbox"/> ₃ | <input type="checkbox"/> ₄ | <input type="checkbox"/> ₅ |
| b) | literacy skills (e.g., reading and writing words) | <input type="checkbox"/> ₁ | <input type="checkbox"/> ₂ | <input type="checkbox"/> ₃ | <input type="checkbox"/> ₄ | <input type="checkbox"/> ₅ |
| c) | mathematical skills (e.g., dealing with numbers and quantities) | <input type="checkbox"/> ₁ | <input type="checkbox"/> ₂ | <input type="checkbox"/> ₃ | <input type="checkbox"/> ₄ | <input type="checkbox"/> ₅ |

| Compared to his/her fellow students, how well do you expect this child to perform at the end of the school year? | | | | | | |
|--|-----------------------------|---------------------------------------|---------------------------------------|---------------------------------------|---------------------------------------|---------------------------------------|
| | far below the class average | below the class average | near the class average | above the class average | far above the class average | |
| a) | in mathematics | <input type="checkbox"/> ₁ | <input type="checkbox"/> ₂ | <input type="checkbox"/> ₃ | <input type="checkbox"/> ₄ | <input type="checkbox"/> ₅ |
| b) | in German language | <input type="checkbox"/> ₁ | <input type="checkbox"/> ₂ | <input type="checkbox"/> ₃ | <input type="checkbox"/> ₄ | <input type="checkbox"/> ₅ |

Appendix C. Statistics on the agreement between rater coding and master coding of videotaped teacher-student interactions

| Variable | Video-rater ID | | | | | |
|---|----------------|-----|-----|-----|-----|-----|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| First training period | | | | | | |
| Content-related agreement (<i>kappa</i>) without predefined subsequences ^a | .66 | – | .46 | .70 | – | .48 |
| Number of subsequently coded school lessons | 6 | 0 | 4 | 1 | 0 | 3 |
| Second training period | | | | | | |
| Content-related agreement (<i>kappa</i>) without predefined subsequences ^a | .69 | .70 | – | – | .62 | .63 |
| Content-related agreement (<i>kappa</i>) with predefined subsequences ^b | .96 | .97 | – | – | .97 | .96 |
| Bivariate correlation between rater coding and master coding regarding ... | | | | | | |
| ... performance relatedness of FB ^c | .86 | .97 | – | – | .93 | .95 |
| ... valence of performance FB | .78 | .92 | – | – | .77 | .81 |
| ... elaborateness of negative performance FB | .70 | .77 | – | – | .78 | .87 |
| Number of subsequently coded school lessons | 17 | 12 | 0 | 0 | 12 | 10 |

Note. Raters 2 and 5 did not participate in the first training period, whereas raters 3 and 4 were no longer available for the second training period.

^a To calculate Cohen's kappa, every interaction sequence was split into 100-msec sections. The resulting estimate indicates the agreement between a rater coding and the master coding, considering differences in the timing of subsequences.

^b This measure is based on content-related ratings of 168 predefined subsequences and indicates the agreement beyond issues of the timing of subsequences.

^c FB = feedback.

References

- Archambault, I., Janosz, M., & Chouinard, R. (2012). Teacher beliefs as predictors of adolescents' cognitive engagement and achievement in mathematics. *The Journal of Educational Research*, 105(5), 319–328. <https://doi.org/10.1080/00220671.2011.629694>.
- 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)]*. IQB – Institut zur Qualitätsentwicklung im Bildungswesen https://doi.org/10.5159/IQB_BIKS_3_10_v4 Version: 4.
- Babad, E. (2009). *The social psychology of the classroom*. New York: Taylor & Francis.
- Bandura, A. (1994). Self-efficacy. In V. S. Ramachandran (Vol. Ed.), *Encyclopedia of human behavior: Vol. 4*, (pp. 71–81). New York: Academic Press.
- 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).
- Baumert, J., & Kunter, M. (2013). The COACTIV model of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.). *Cognitive activation in the mathematics classroom and professional competence of teachers* (pp. 25–48). New York: Springer Science + Business Media.
- Baumert, J., & Maaz, K. (2006). Das theoretische und methodische Konzept von PISA zur Erfassung sozialer und kultureller Ressourcen der Herkunftsfamilie: Internationale und nationale Rahmenkonzeption [The theoretical and methodological concept of PISA to assess social and cultural resources of the family of origin: International and national conceptual framework]. In J. Baumert, P. Stanat, & R. Waterman (Eds.). *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Becker, D. (2013). The impact of teachers' expectations on students' educational opportunities in the life course: An empirical test of a subjective expected utility explanation. *Rationality and Society*, 25(4), 422–469. <https://doi.org/10.1177/1043463113504448>.
- 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(1), 168–179. <https://doi.org/10.1037/a0017289>.
- de Boer, H., & van der Werf, M. P. C. (2015). Influence of misaligned parents' aspirations on long-term student academic performance. *Educational Research and Evaluation*, 21(3), 232–257. <https://doi.org/10.1080/13803611.2015.1039548>.
- Bohn, C. M., Roehrig, A. D., & Pressley, M. (2004). The first days of school in the classrooms of two more effective and four less effective primary-grades teachers. *The Elementary School Journal*, 104(4), 269–287. <https://doi.org/10.1086/499753>.
- van Buuren, S., Brand, J. P. L., Groothuis-Oudshoorn, C. G. M., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*, 76(12), 1049–1064. <https://doi.org/10.1080/10629360600810434>.
- Chen, Y.-H., Thompson, M. S., Kromrey, J. D., & Chang, G. H. (2011). Relations of student perceptions of teacher oral feedback with teacher expectancies and student self-concept. *The Journal of Experimental Education*, 79(4), 452–477. <https://doi.org/10.1080/00220973.2010.547888>.
- Chow, S. L. (1987). Some reflections on Harris and Rosenthal's 31 meta-analyses. *Journal of Psychology*, 121(1), 95–100. <https://doi.org/10.1080/00223980.1987.9712648>.
- Clarke, S., Timperley, H. S., & Hattie, J. (2003). *Unlocking formative assessment: Practical strategies for enhancing students' learning in the primary and intermediate classroom*. Auckland, New Zealand: Hodder Moa Beckett.
- Dean, C. B., Hubbell, E. R., Pitler, H., & Stone, B. (2012). *Classroom instruction that works: Research-based strategies for increasing student achievement* (2nd ed.). Denver: McREL.
- Deutscher, T., & Selter, C. (2013). Frühe mathematische Bildung – forschungsbefunde und Förderkonzepte [Early mathematical education—Empirical evidence and training approaches]. In M. Stamm, & D. Edelman (Eds.). *Handbuch frühkindliche bildungsforschung* (pp. 543–556). Wiesbaden: Springer VS.
- Dietrich, J., Dicke, A.-L., Kracke, B., & Noack, P. (2015). Teacher support and its influence on students' intrinsic value and effort: Dimensional comparison effects across subjects. *Learning and Instruction*, 39, 45–54. <https://doi.org/10.1016/j.learninstruc.2015.05.007>.
- Dijksterhuis, A., & Bargh, J. A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. *Advances in Experimental Social Psychology*, 33, 1–38.
- Fazio, R. H., & Olson, M. A. (2014). The MODE model: Attitude-behavior processes as a function of motivation and opportunity. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.). *Dual process theories of the social mind* (pp. 155–171). New York: Guilford Press.
- 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. <https://doi.org/10.1016/j.cedpsych.2014.10.006>.
- Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102(4), 652–670.
- 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*. Lisbon, Portugal: Annual Conference of the International Social Survey Programme Paper presented at the.
- GANZEBOOM, H. B. G., GRAAF, P. M. D., & TREIMAN, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21, 1–56.
- Gentrup, S., & Rjosk, C. (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. <https://doi.org/10.1080/13803611.2018.1550840>.
- 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. <https://doi.org/10.1007/s11618-018-0806-2>.
- Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, 102, 74–78. <https://doi.org/10.1016/j.paid.2016.06.069>.
- Gill, S., & Reynolds, A. J. (1999). Educational expectations and school achievement of urban African American children. *Journal of School Psychology*, 37(4), 403–424. [https://doi.org/10.1016/S0022-4405\(99\)00027-8](https://doi.org/10.1016/S0022-4405(99)00027-8).
- Harris, M., & Rosenthal, R. (1985). Mediation of interpersonal expectancy effects: 31 meta-analyses. *Psychological Bulletin*, 97(3), 363–386. [0033-2909/95/3.363](https://doi.org/10.1037/0033-2909.95.3.363).
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>.
- 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(3), 662–670. <https://doi.org/10.1037/a0014306>.
- Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement:

- A review of literature. *Review of Educational Research*, 59(3), 297–313. <https://doi.org/10.3102/00346543059003297>.
- 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. <https://doi.org/10.1007/s11218-017-9384-z>.
- Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In L. M. Le Cam, & J. Neyman (Vol. Eds.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability: Vol. 1*, (pp. 221–233). Berkeley: University of California Press.
- Hurwitz, J. T., Elliott, S. N., & Braden, J. P. (2007). The influence of test familiarity and student disability status upon teachers' judgments of students' test performance. *School Psychology Quarterly*, 22(2), 115–144.
- Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, 73(2), 509–527.
- Jenkins, L. N., & Demaray, M. K. (2016). Teachers' judgments of the academic achievement of children with and without characteristics of inattention, impulsivity, and hyperactivity. *Contemporary School Psychology*, 20(2), 183–191. <https://doi.org/10.1007/s40688-015-0073-7>.
- Johnson, T. J., Feigenbaum, R., & Weiby, M. (1964). Some determinants and consequences of the teacher's perception of causation. *Journal of Educational Psychology*, 55(5), 237–246. <https://doi.org/10.1037/h0043389>.
- Jose, P. E. (2013). *Doing statistical mediation and moderation*. London: Guilford Press.
- 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.
- Jungbluth, P. (1994). Lehrererwartungen und Ethnizität. Innerschulische Chancendeterminanten bei Migrantenschülern in den Niederlanden [Teacher expectations and ethnicity. The educational opportunities of adolescent migrants in The Netherlands]. *Zeitschrift für Pädagogik*, 40(1), 113–125 doi:urn:nbn:de:0111-pedocs-108314.
- Juska-Bacher, B. (2013). Leserelevante Kompetenzen und ihre frühe Förderung [Reading-related competencies and their early training]. In M. Stamm, & D. Edelmann (Eds.). *Handbuch frühkindliche bildungsforschung* (pp. 485–500). Wiesbaden: Springer VS.
- Jussim, L. (1986). Self-fulfilling prophecies: A theoretical and integrative review. *Psychological Review*, 93(4), 429–445. <https://doi.org/10.1037/0033-295X.93.4.429>.
- 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 (Vol. Ed.), *Advances in experimental social psychology: Vol.28*, (pp. 281–388). San Diego: Academic Press.
- Jussim, L., & Harber, K. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review*, 9(2), 131–155. https://doi.org/10.1207/s15327957pspr0902_3.
- Jussim, L., Robustelli, S., & Cain, T. (2009). Teacher expectations and self-fulfilling prophecies. In K. Wentzel, & A. Wigfield (Eds.). *Handbook of motivation at school* (pp. 349–379). New York: Routledge.
- Kristen, C., Stanat, P., Kogan, I., Lorenz, G., Gentrup, S., & Rahmann, S. (2018a). *Kompetenzerwerb und lernvoraussetzungen (KuL) [competence acquisition and learning preconditions (KuL)]*. IQB – Institut zur Qualitätsentwicklung im Bildungswesen https://doi.org/10.5159/IQB_KuL_v1 Version: 1.
- Kristen, C., Stanat, P., Kogan, I., Lorenz, G., Gentrup, S., & Rahmann, S. (2018b). *Kompetenzerwerb und lernvoraussetzungen (KuL) [competence acquisition and learning preconditions (KuL)]*. Forschungsdatenzentrum Bildung am DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation. Observational data <https://doi.org/10.7477/287:1:0>.
- Kuklinski, M. R., & Weinstein, R. S. (2001). Classroom and developmental differences in a path model of teacher expectancy effects. *Child Development*, 72, 1554–1578.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>.
- 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. (2019). *Subtle discrimination? Stereotypes among teachers trigger ethnic bias in their achievement expectations but are unrelated to ethnic achievement gaps*. (in preparation).
- 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(1), 89–111. <https://doi.org/10.1007/s11577-015-0352-3>.
- Ma, X. (2001). Participation in advanced mathematics: Do expectation and influence of students, peers, teachers, and parents matter? *Contemporary Educational Psychology*, 26(1), 132–146. <https://doi.org/10.1006/ceps.2000.1050>.
- Madon, S., Jussim, L., & Eccles, J. (1997). In search of the powerful self-fulfilling prophecy. *Journal of Personality and Social Psychology*, 72, 791–809. <https://doi.org/10.1037/0022-3514.72.4.791>.
- van Matre, J. C., Valentine, J. C., & Cooper, H. M. (2000). Effect of students' after-school activities on teachers' academic expectancies. *Contemporary Educational Psychology*, 25, 167–183. <https://doi.org/10.1006/ceps.1998.0999>.
- 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 Psychology*, 32(1), 159–184. <https://doi.org/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(3), 235–261. <https://doi.org/10.1016/j.jsp.2007.05.001>.
- Mehmetoglu, M. (2018). medsem: a Stata package for statistical mediation analysis. *International Journal of Computational Economics and Econometrics*, 8(1), 63–78.
- Meissel, K., Meyer, F., Yao, E. S., & Rubie-Davies, C. M. (2017). Subjectivity of teacher judgments: Exploring student characteristics that influence teacher judgments of student ability. *Teaching and Teacher Education*, 65, 48–60. <https://doi.org/10.1016/j.tate.2017.02.021>.
- Merton, R. K. (1948). The self-fulfilling prophecy. *Antioch Review*, 8(2), 193–210. <https://doi.org/10.2307/4609267>.
- Morris, E. W. (2005). From “middle class” to “trailer trash:” Teachers' perceptions of white students in a predominately minority school. *Sociology of Education*, 78(2), 99–121. <https://doi.org/10.1177/003804070507800201>.
- Muntoni, F., & Retelsdorf, J. (2018). Gender-specific teacher expectations in reading—the role of teachers' gender stereotypes. *Contemporary Educational Psychology*, 54, 212–220. <https://doi.org/10.1016/j.cedpsych.2018.06.012>.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220.
- OECD (2003). *Literacy skills for the world of tomorrow: Further results from PISA 2000*. Paris: OECD Publishing.
- Pellegrini, R. J., & Hicks, R. A. (1972). Prophecy effects and tutorial instruction for the disadvantaged child. *American Educational Research Journal*, 3, 413–419. <https://doi.org/10.2307/1161757>.
- Pendry, L. (2015). Social cognition. In M. Hewstone, W. Stroebe, & K. Jonas (Eds.). *An introduction to social psychology* (pp. 93–122). (6 ed.). Chichester: Wiley.
- Peterson, E. R., Rubie-Davies, C. M., Osborne, D., & Sibley, C. (2016). Teachers' explicit expectations and implicit prejudiced attitudes to educational achievement: Relations with student achievement and the ethnic achievement gap. *Learning and Instruction*, 42, 123–140.
- 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 GmbH.
- Ready, D. D., & Chu, E. M. (2015). Sociodemographic inequality in early literacy development: The role of teacher perceptual accuracy. *Early Education & Development*, 26(7), 970–987. <https://doi.org/10.1080/10409289.2015.1004516>.
- 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(2), 335–360. <https://doi.org/10.3102/0002831210374874>.
- Richter, D., Kuhl, P., & Pant, H. A. (2012). Soziale disparitäten [social disparities]. In P. Stanat, H. A. Pant, K. Böhme, & D. Richter (Eds.). *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (pp. 191–207). Münster: Waxmann.
- Riegle-Crumb, C., & Humphries, M. (2012). Exploring bias in math teachers' perceptions of students' ability by gender and race/ethnicity. *Gender & Society*, 26(2), 290–322. <https://doi.org/10.1177/0891243211434614>.
- Roorda, D. L., Koomen, H. M. Y., Spilt, J. L., & Oort, F. J. (2011). The influence of affective teacher–student relationships on students' school engagement and achievement: A meta-analytic approach. *Review of Educational Research*, 81(4), 493–529. <https://doi.org/10.3102/0034654311421793>.
- Rosenthal, R. (1973). On the mediation of pygmalion effects: A four factor “theory” *Papua New Guinea Journal of Education*, 9, 1–12.
- Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*, 16–20.
- Rubie-Davies, C. M. (2007). Classroom interactions: Exploring the practices of high- and low-expectation teachers. *British Journal of Educational Psychology*, 77, 289–306. <https://doi.org/10.1348/000709906X101601>.
- Rubie-Davies, C. M., Hattie, J. A. C., Townsend, M. A. R., & Hamilton, R. J. (2007). Aiming high: Teachers and their students. In V. N. Galwey (Ed.). *Progress in educational psychology research* (pp. 65–91). Hauppauge, NY: Nova Publisher.
- Rubie-Davies, C. M., & Peterson, E. R. (2011). Teacher expectations and beliefs. Influences on the socioemotional environment of the classroom. In C. M. Rubie-Davies (Ed.). *Educational psychology. Concepts, research and challenges* (pp. 135–149). New York: Routledge.
- Rubie-Davies, C. M., & Peterson, E. R. (2016). Relations between teachers' achievement, over- and underestimation, and students' beliefs for Māori and Pākehā students. *Contemporary Educational Psychology*, 47, 72–83. <https://doi.org/10.1016/j.cedpsych.2016.01.001>.
- Rubie-Davies, C. M., Peterson, E. R., Sibley, C. G., & Rosenthal, R. (2015). A teacher expectation intervention: Modelling the practices of high expectation teachers. *Contemporary Educational Psychology*, 40, 72–85. <https://doi.org/10.1016/j.cedpsych.2014.03.003>.
- Rubie-Davies, C. M., Weinstein, R. S., Huang, F. L., Gregory, A., Cowan, P. A., & Cowan, C. P. (2014). Successive teacher expectation effects across the early school years. *Journal of Applied Developmental Psychology*, 35(3), 181–191. <https://doi.org/10.1016/j.appdev.2014.03.006>.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: John Wiley & Sons.
- Schenke, K., Nguyen, T., Watts, T. W., Sarama, J., & Clements, D. H. (2017). Differential effects of the classroom on african american and non-african american's mathematics achievement. *Journal of Educational Psychology*. <https://doi.org/10.1037/edu0000165>.
- Schuchardt, C., & Dunkake, I. (2014). Schichtspezifische Stereotype unter angehenden Lehrkräften [Social class stereotypes among teacher trainees]. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, 34(1), 89–107. <https://doi.org/10.3262/ZSE1401089>.
- Shifer, D. (2013). Stigma of a label: Educational expectations for high school students labeled with learning disabilities. *Journal of Health and Social Behavior*, 54(4),

- 462–480. <https://doi.org/10.1177/0022146513503346>.
- Shifrer, D. (2016). Stigma and stratification limiting the math course progression of adolescents labeled with a learning disability. *Learning and Instruction*, 42, 47–57. <https://doi.org/10.1016/j.learninstruc.2015.12.001>.
- Snow, R. E. (1969). Review of "pygmalion in the classroom" by rosenthal and jacobsen. *Current Directions in Psychological Science*, 14, 197–199.
- Spinath, B., & Steinmayr, R. (2008). Longitudinal analysis of intrinsic motivation and competence beliefs: Is there a relation over time? *Child Development*, 79(5), 1555–1569.
- Stahl, N. (2007). Schülerwahrnehmung und -beurteilung durch Lehrkräfte [Student perception and assessment by teachers]. In H. Ditton (Ed.). *Kompetenzaufbau und Laufbahnen im Schulsystem* (pp. 171–198). Münster: Waxmann.
- StataCorp, L. L. C. (1985-2017). *Stata/SE 15.1 [computer software]*. Texas: College Station.
- StataCorp (2017). *Stata user's guide. Release, Vol. 15*. Texas: Stata Press.
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762. <https://doi.org/10.1037/a0027627>.
- Thorndike, R. L. (1968). Review of robert rosenthal and lenore Jacobson, pygmalion in the classroom. *American Educational Research Journal*, 5(4), 708–711. <https://doi.org/10.2307/1162010>.
- 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(2), 217–240. <https://doi.org/10.1007/s11218-015-9326-6>.
- Timmermans, A. C., Kuyper, H., & van der Werf, G. (2015). Accurate, inaccurate, or biased teacher expectations: Do Dutch teachers differ in their expectations at the end of primary education? *British Journal of Educational Psychology*, 85(4), 459–478. <https://doi.org/10.1111/bjep.12087>.
- Urhahne, D. (2015). Teacher behavior as a mediator of the relationship between teacher judgment and students' motivation and emotion. *Teaching and Teacher Education*, 45, 73–82. <https://doi.org/10.1016/j.tate.2014.09.006>.
- Vogt, F., & Rogalla, M. (2009). Developing adaptive teaching competency through coaching. *Teaching and Teacher Education*, 25, 1051–1060. <https://doi.org/10.1016/j.tate.2009.04.002>.
- Wang, S., Rubie-Davies, C. M., & Meissel, K. (2018). A systematic review of the teacher expectation literature over the past 30 years. *Educational Research and Evaluation*, 24(3–5), 124–179. <https://doi.org/10.1080/13803611.2018.1548798>.
- Weinstein, R. S. (1983). Student perceptions of schooling. *The Elementary School Journal*, 83(4), 286–312.
- Weiß, R. H., & Osterland, J. (1997). *Grundintelligenztest skala 1: CFT 1. Handanweisung [culture fair intelligence test 1: CFT1. Manual] (5., revidierte auflage ed.)*. Göttingen: Hogrefe.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), 817–838.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1), 1–25.
- Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering baron and kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, 37, 197–206. <https://doi.org/10.1086/651257>.
- Zhou, J., & Urhahne, D. (2013). Teacher judgment, student motivation, and the mediating effect of attributions. *European Journal of Psychology of Education*, 28(2), 275–295.