



# Exploring teacher popularity: associations with teacher characteristics and student outcomes in primary school

Benjamin Fauth<sup>1,2</sup>  · Jasmin Decristan<sup>2,3</sup> · Svenja Rieser<sup>3</sup> · Eckhard Klieme<sup>2,4</sup> · Gerhard Büttner<sup>2,5</sup>

Received: 31 October 2017 / Accepted: 26 June 2018 / Published online: 22 August 2018  
© The Author(s) 2018

## Abstract

In this study, we investigated conditions and consequences of teacher popularity in primary schools. Teacher popularity is embedded in a theoretical framework that describes relationships between teacher competence, teaching quality, and student outcomes. In the empirical analyses, we used multilevel modeling to distinguish between individual students' liking of the teacher and a teacher's popularity as rated by the whole class ( $N=1070$  students, 54 teachers). The classroom level composite of the extent to which students liked their teacher was a reliable indicator of teacher popularity. Teacher popularity was associated with teacher self-reports of self-efficacy and teaching enthusiasm and with external observers' ratings of teaching quality. The grades students received were not related to the popularity ratings. In a longitudinal study, teacher popularity predicted students' learning gains and interest development over and above the effects of teaching quality. These results suggest that teacher popularity can be a useful and informative indicator in research on students' academic development and teacher effectiveness.

**Keywords** Teacher popularity · Teaching quality · Enthusiasm · Self-efficacy · Primary school

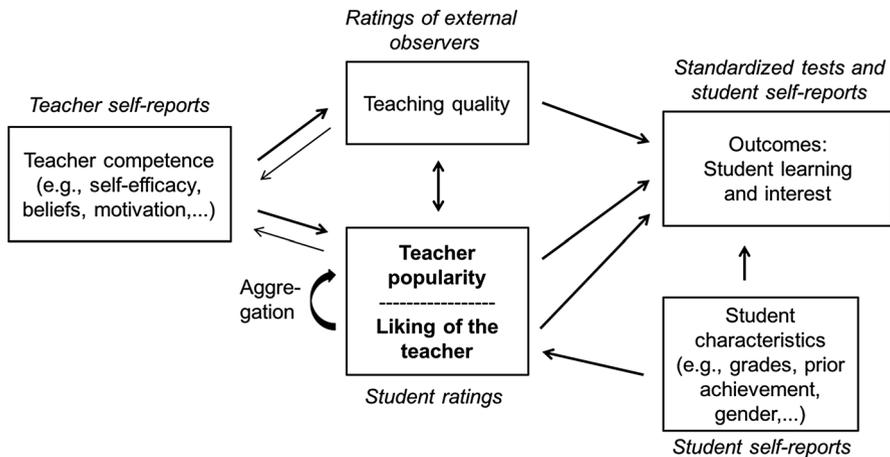
## 1 Introduction

The popularity of teachers is an issue that is often discussed but seldom conceptualized as a construct in a clear way. An essential feature of a person's popularity is that (s)he is liked by more than one other person. Thus, popularity is distinguished from individual liking. At school, every student can report which teachers (s)he likes more or less than others. Students' reasons for liking a teacher (or not) can be diverse. However, students in the same classroom can be expected to exhibit some agreement about how

---

✉ Benjamin Fauth  
benjamin.fauth@uni-tuebingen.de

Extended author information available on the last page of the article



**Fig. 1** Theoretically proposed relations between teacher popularity, teacher competence, and teaching quality (modified model according to Kunter et al. 2013). *Note* Preferred sources of data for the respective constructs are in italics

much they like a certain teacher. Accordingly, in every school, teachers tend to show variability in how popular they are. In the present contribution, we tried to disentangle these two aspects—the individual student’s liking of a teacher (as a characteristic of the individual student–teacher relationship) and the popularity of a teacher in a classroom (as an ascribed characteristic of the teacher).

The analyses presented in this contribution were motivated by the assumption that teacher popularity can be a useful and informative indicator in research on students’ academic development and teacher effectiveness. However, although researchers have touched upon the topic of teacher popularity, no previous studies have provided answers for how the concept of teacher popularity can be theoretically conceptualized and empirically assessed. To take a first step in this direction, we examined how teacher popularity can be embedded in an established and widely acknowledged model of teacher competence, including teaching quality as an additional component that can contribute to a deeper understanding of classroom processes (Baumert and Kunter 2013; Klieme et al. 2009; Kunter et al. 2013; see below and Fig. 1). In a second step, we empirically tested whether teacher popularity matters for student development: Are popular teachers more successful in promoting students’ conceptual understanding and the development of subject-related interest? To make sure that these effects are not merely effects of the higher teaching quality of popular teachers, we tested effects of teacher popularity on student outcomes beyond what is explained by teaching quality.

## 2 Theoretical background

### 2.1 How can teacher popularity be conceptualized and assessed?

Previous research on the concept of teacher popularity has primarily concentrated on its devaluating relation to student ratings of teaching quality. Aleamoni (1999) summarized the concerns that are typically expressed by researchers: “Most student rating schemes are nothing more than a popularity contest with the warm, friendly, humorous instructor emerging as the winner every time” (p. 154). The ultimate concern is that a teacher’s popularity has nothing to do with the actual quality of instruction or professional competence. Such concerns have also been tested in research on the “Dr. Fox effect” (Marsh 1987), which refers to a correlation between teacher expressiveness and student evaluations, independent of the content of a lecture. However, discussions about the Dr. Fox effect have shown that factors that were previously considered to represent bias may indeed contain valid and informative variance (Marsh and Ware 1982). As far back as 1938, Corey and Beery stated that teacher popularity “might be one valid criterion of teaching effectiveness. Liking a teacher might be closely related to learning” (Corey and Beery 1938, p. 665).

In the present contribution, we follow this idea by asking whether teacher popularity offers interesting and potentially relevant information that can help explain students’ progress in school. We define teacher popularity as students’ affectively tinged, shared general impression of their teacher (Atamian and Gan-guli 1993; Fauth et al. 2014a; Payne 1987; Wagner 2008). This definition has several implications: 1. Teacher popularity is a teacher characteristic that varies (i.e., teachers can be more or less popular). 2. It is a global construct that does not refer to specific distinct traits but to the teacher as a whole person. 3. Popularity is shaped by interactions between teachers and students. 4. The appropriate data source for measuring popularity is student ratings. Class-wide student ratings are constitutive for assessing teacher popularity in that class. 5. Popularity can be affected by personal characteristics of the teacher, by teaching behavior in the classroom (teaching quality), and by characteristics of the students who serve as raters of popularity.

Following these implications, it is obvious that the students in a certain class play an important role in determining teacher popularity: The popularity of a teacher depends on the class (s)he teaches, and teacher popularity cannot be rated from the outside (e.g., via video-based observations). A simple but convincing operationalization would involve using student survey items such as “I like my teacher” (Gruehn 2000; Wagner 2008). To obtain an indicator of teacher popularity, individual students’ liking of the teacher are then aggregated at the classroom level (see below). Such items are global as they do not specify which particular aspects of the teacher’s personality or teaching behavior students should base their judgments on. In addition, they tend to capture a more emotional evaluation of the teacher rather than rational judgments that are based on objective criteria

of a teacher's behavior in the classroom. These kinds of items are very well suited to measure the teacher popularity construct defined above.

The use of such items in class-wide assessments can tap two kinds of information. The first kind of information represents the students' liking of the teacher as a characteristic of individual student–teacher relationships. A student's appreciation of the teacher and bonding between the student and teacher will be relevant for the development of these relationships (Davis 2003). Second, by aggregating the extent to which individual students like the teacher, we can obtain a measure of a teacher's popularity. If the students in a classroom have at least some agreement in the extent to which they like the teacher, the class-averaged scores representing the liking of a teacher can be considered a teacher characteristic. However, teacher popularity is considered an ascribed teacher characteristic, i.e. it is actually not independent of the students (which would be the case for teachers' intelligence or other personality traits).

## 2.2 Embedding teacher popularity in a broader theoretical framework

We embedded the concept of teacher popularity in a model of teacher competence and teaching quality that originated from the German COACTIV study, a longitudinal extension to the PISA 2003 assessment that has gained much attention in recent years (Kunter et al. 2013). In the theoretical foundations of this study, the researchers distinguished characteristics of the teacher from characteristics of the teacher's actual teaching in the classroom and student outcomes as the product of teaching. The personal characteristics have also been referred to as teachers' *professional competence*. In this area, researchers have paid special attention to different aspects of teaching-specific personal characteristics of teachers, such as professional knowledge, professional beliefs about teaching and learning, and motivation to face the challenges of everyday classroom instruction (Baumert and Kunter 2013). In contrast to these teacher characteristics, *teaching quality* refers to the quality of instruction that a teacher is able to implement while interacting with the students in the classroom (Klieme et al. 2009). In this area, three basic dimensions of teaching quality have convincingly been described: cognitive activation, supportive climate, and classroom management (Fauth et al. 2014a, b; Klieme et al. 2009). In its simplest form, the model suggests that teacher competence positively affects teaching quality, which, in turn, has a positive impact on *students' academic development* (the upper part of the model presented in Fig. 1).

According to our conceptualization, teacher popularity is related to teachers' professional competence and to characteristics of teaching quality in several ways. A teacher's enthusiasm may have an effect on how popular the teacher is. Conversely, a teacher's popularity may affect her/his motivation for teaching. With regard to teaching quality, the way a teacher supports the students may have an impact on her/his popularity, and a teacher's respect for the students may be influenced by her/his popularity. Most important, a substantial relationship is assumed between teacher popularity and student learning, which significantly contributes to the prediction of student outcomes over and above teacher competence and teaching quality. The

way in which we operationalize teacher popularity does not refer to specific teacher behaviors in the classroom as the teaching quality measures do (see Sect. 2.3.2). Also, teacher popularity is not independent from the class a teacher teaches as it is assumed to be the case for the measures of teachers' professional competence.

In the present investigation, we explored some of the theoretically assumed relations between teacher popularity on one side and teacher competence and teaching quality on the other. As teacher popularity is conceptualized as rated by students, we also included student characteristics as an antecedent of teacher popularity in the model (Fig. 1). We thus examined (1) what teacher competence and teaching quality can contribute to the explanation of teacher popularity and (2) whether teacher popularity can contribute to explaining student learning over and above teaching quality. In doing so, we investigated (1) the antecedents of teacher popularity and (2) whether teacher popularity can provide relevant information for explaining student learning beyond the constructs that are usually investigated in teacher effectiveness research. As mentioned above, students' ratings can be relevant at the individual and at the classroom level of analysis. The analyses based on this model might reveal that teacher popularity is not only a factor of bias in student ratings, but it is also an informative indicator of teacher effectiveness that provides additional information beyond the constructs that are usually considered when it comes to the prediction of student outcomes.

All of the constructs presented in Fig. 1 can play different roles depending on grade levels and student ages. This holds particularly true for the constructs of liking the teacher and teacher popularity. The relationships between teachers and students look different at different stages of schooling. In primary school for example, teachers do not only act as instructors, but they also play an important role as caregivers. Primary school students have different needs regarding instruction as well as regarding social and emotional issues compared to students at the secondary or university level (Eccles et al. 1993). We believe that interpersonal factors such as students' liking of the teacher and teacher popularity play a particularly important role in primary school education.

## 2.3 What makes a teacher popular in the eyes of students?

The model presented in Fig. 1 indicates that there are three groups of factors that are associated with teacher popularity: personal characteristics of the teacher, teaching quality, and individual characteristics of the students.

### 2.3.1 Teacher characteristics

Professional competence refers to a set of knowledge areas, beliefs, and motivational variables that are empirically connected to teachers' vocational success in terms of teaching quality and student learning gains (Baumert and Kunter 2013). Research on the Dr. Fox effect indicates that teachers' motivation and beliefs are especially crucial for a teacher's popularity.

Among the motivational variables, teacher *enthusiasm* has earned special attention. This concept also plays a role in discussions of the Dr. Fox effect and teacher popularity mentioned above, and this makes the concept interesting for the present investigation. Kunter et al. (2008) distinguished between two dimensions of teacher enthusiasm: enthusiasm for the subject and enthusiasm for teaching. Empirical results have shown that only *enthusiasm for teaching* is relevant for students' perceptions of teaching quality and student outcomes (Feldman 1986; Kunter et al. 2008, 2011).

For teachers' beliefs, we can distinguish beliefs concerning the self (e.g., *self-efficacy*) from beliefs about the nature of teaching and learning (e.g., *constructivist beliefs*). Self-efficacy is conceptualized as teachers' self-perceptions of competence to perform well in their job including the management of potentially challenging situations in everyday school practice (Bandura 1995; Guo et al. 2014; Schmitz and Schwarzer 2000; Zee et al. 2016). The concept of constructivist beliefs is based on the assumption that teaching is more a promotion of individual knowledge construction rather than the direct transmission of knowledge from teacher to student (Dubberke et al. 2008; Staub and Stern 2002). Although both kinds of beliefs seem important for teaching, they tap into very different aspects of teacher competence. With regard to teacher popularity, self-efficacy seems especially important as students should value a teacher with a self-confident appearance in the classroom (Lui and Bonner 2016). By contrast, research has yet to determine whether students are sensitive to teachers' subject-related beliefs about the nature of teaching and learning such as constructivist beliefs.

In summary, three concepts of professional competence have to be further investigated in their relation to teacher popularity: enthusiasm for teaching, self-efficacy, and constructivist beliefs. Previous research on teacher competence has identified these aspects as key variables in the area of teacher motivation and beliefs (Holzberger et al. 2013; Kunter et al. 2008).

### 2.3.2 Classroom process features: teaching quality

In his seminal quote, Aleamoni (1999) mentioned warmth, friendliness, and humour as features that make a teacher the "winner every time" in student evaluations. However, similar teacher behavior has also been conceptualized as part of classroom process quality or *teaching quality*. For instance, Klieme et al. (2009) integrated friendliness, respect for the students, and individual learning support into the concept of "supportive climate". Supportive climate is regarded as one of three basic teaching quality dimensions that have become very famous in German-speaking countries in recent years. The other two dimensions are cognitive activation and classroom management. Classroom management refers to teachers' strategies to prevent disruptions and keep order in the classroom with the aim of maximizing students' time on task. Cognitive activation refers to subject-specific strategies to foster students' cognitive engagement using complex and challenging tasks, exploring students' prior concepts and ideas, and the practice of the Socratic Dialogue (Baumert et al. 2010; Lipowsky et al. 2009). This three-dimensional framework is very similar to the domains covered in the Classroom Assessment Scoring System (CLASS) framework (Pianta and

Hamre 2009). In particular, the dimensions of cognitive activation and classroom management have been shown to predict student learning (Kyriakides et al. 2013; Lipowsky et al. 2009), whereas supportive climate was found to be especially connected to students' motivational and interest development (Kunter et al. 2013).

Wagner (2008) and Fauth et al. (2014a) reported substantial correlations between student ratings of teaching quality and teacher popularity. Nonetheless, the two concepts formed distinct dimensions in confirmatory factor analyses (Fauth et al. 2014a). Thus, student ratings of teaching quality cannot be reduced to teacher popularity. The highest correlations with teacher popularity were found for measures of supportive climate (e.g., student orientation in Wagner 2008), and considerably lower correlations were found for classroom management features. These are comprehensible relations as supportive climate also refers to positive teacher-student relationships (Lee 2012). However, these studies used only student ratings to tap teaching quality. In the present study, we used video-based observations of teaching quality to avoid obtaining flawed correlations as a result of common method bias.

### 2.3.3 Student characteristics

The aforementioned characteristics of teacher competence and teaching quality might be related to teacher popularity at the classroom level (Fig. 1). When it comes to the extent to which an individual student likes the teacher, associations with individual characteristics of the students become relevant. Here, the question is: Which students tend to rate their individual relationship with the teacher in a more positive manner?

It seems that girls tend to give more positive ratings of teaching quality (Fauth et al. 2016; Benton and Cashin 2012; Wagner 2008, p. 117). Research has yet to determine whether this tendency is grounded in a leniency/severity bias or reflects valid experiences of the students in a class. In addition to their more positive ratings of teaching quality, girls also report *liking* their teachers more than boys do (Wagner 2008, p. 116).

Another individual student characteristic that has been widely discussed to affect students' ratings of their teachers is *grading leniency* (Benton and Cashin 2012). According to the grading leniency hypothesis, teachers who assign better grades than a student deserves will be more popular with that student. In a meta-analysis, Centra (2003) reported correlations of .10 to .30 between expected grades and students' ratings of teaching quality. In the present investigation, we expected connections between grades and the extent to which students liked the teacher.

## 2.4 Consequences of liking the teacher and teacher popularity: effects on student outcomes

Meta-analyses have shown that positive student-teacher relationships affect student learning and motivation (e.g., Cornelius-White 2007). Beyond these results, research on the effects of teachers' popularity is scarce. Montalvo et al. (2007) asked high-school students to think about one current teacher they liked a lot and

to complete a survey “as it relates to that teacher and the class he/she teaches” (p. 147). The same participants were asked to think about a teacher they disliked a lot and to again complete the survey, this time with regard to the disliked teacher. Their results revealed differences in students’ self-reported motivation, effort, and semester grades—with higher scores for classes taught by teachers the students liked. Montalvo et al. (2007) claimed that these effects could be explained by the concept of “pleasing the teacher”. According to this idea, students put more effort into learning in classes taught by teachers they like because they do not want to disappoint the teacher with a poor performance. It is clear that motivation and intrinsic interest play an important role in the interpretation of these results (Deci and Ryan 2000; Wigfield and Eccles 2000). Montalvo et al. (2007) suggested that popular teachers are able to influence students’ subject-related value beliefs and the instrumentality of schoolwork for future goals.

Montalvo et al.’s (2007) results provided interesting insights, but their exploratory power for the primary school context that the current study focuses on was limited. In addition, one limitation of the above-described research is that these studies were not able to distinguish between different levels of analysis (the classroom and individual levels). Thus, it is unclear whether the reported effects should be regarded as teacher effects (popular teachers receive more “likes” and produce better results) or as an effect of the specific student–teacher relationship (a student likes a teacher—and may want to please this teacher—which in turn leads to better results for that student compared with another one).

A major goal of the present investigation was to disentangle these effects by applying multilevel modeling. As mentioned above, in students’ reports of how much they like the teacher, two sources of variance can be considered: the ratings of individual (idiosyncratic) students and the (shared) ratings of the students in the class (Lüdtke et al. 2009). The former refers to the individual relationship between a student and his or her teacher and is reflected by variance *within* classes. The latter refers to the popularity of a teacher as rated by an entire class of students and is reflected by variance *between* classes. A prerequisite for these multilevel analyses is a substantial amount of variance between classes (i.e., agreement between students in the same class concerning how much they like their teacher; Lüdtke et al. 2009), which then forms the construct of teacher popularity. This aggregation follows a fuzzy composition process as described in the framework from Bliese (2000), meaning that the constructs at the two levels of analysis (students’ individual liking of the teacher and teacher popularity) are related to each other but not the same construct.

Applying such multilevel analyses, Wagner (2008) reported cross-sectional associations between a single-item measure (“I like my teacher”) and measures of achievement (standardized tests and grades) in a large secondary school sample from Germany. Bivariate correlations showed significant relations between students’ liking of the teacher and achievement (within classes) but no correlations between teacher popularity and achievement (between classes).

As mentioned above, Wagner (2008) also reported high correlations between teacher popularity and measures of teaching quality. As teaching quality should also affect student learning and motivation, it was especially important for the present investigation to examine the effects of teacher popularity on student interest and learning over and

above the effects of classroom process quality. Is there something about popular teachers that promotes student learning and achievement beyond what is explained by the quality of their teaching? We assume that this question will be particularly important in primary schools where interpersonal relations between a teacher and the class are a crucial factor for students' academic development (Pianta and Hamre 2009).

### 3 Research questions

Several research questions were inspired from the aforementioned empirical results and theoretical considerations.

- (1) Is there a “teacher popularity” construct that can be reliably assessed via student ratings at the classroom level of analysis in primary schools?
- (2) What are the relations between teacher popularity and (a) teacher enthusiasm, constructivist beliefs, and self-efficacy, (b) students' gender and grades, and (c) the basic dimensions of teaching quality as rated by external observers?
- (3) What are the effects of teacher popularity on the development of student achievement and students' subject-related interest? (a) Is teacher popularity a relevant predictor of these student outcomes? (b) Do these effects refer to the individual level or to the classroom level of analysis? (c) Do these effects exist over and above the quality of teaching?

## 4 Method

### 4.1 Sample

Our analyses drew on longitudinal data from 1070 third-grade students and their 54 science teachers (project IGEL, Decristan et al. 2015; Hardy et al. 2011). These students participated in an intervention study that followed the design described in Sect. 4.3. The average student age was 8.8 years ( $SD=0.50$ ), and 49% of the students were female. Participating teachers had a mean age of 42.8 years ( $SD=9.2$ ) and professional experiences of 16.4 years on average ( $SD=8.6$  years). They taught science education in the participating classes. The target populations of the study were students and teachers from public primary schools in a German state. Participating schools were located in both urban (61% of classes) and rural areas. Participation in the study was voluntary for both teachers and students. Teachers gave their informed consent to participate in data collection. Parents gave their informed consent for students' participation. The average participation rate for each classroom was 96%. The data collection in this study was approved by the Ethics Committee of the Faculty of Psychology and Sports Sciences at Goethe-University Frankfurt, Germany.

## 4.2 Instruments

We used different sources of data to capture the constructs we were interested in: student surveys and standardized achievement tests, teacher self-report surveys, and standardized video observations in the classroom.

### 4.2.1 Student questionnaires

Teacher popularity was measured with a three-item scale based on Wagner (2008; “I like my teacher very much”, “My teacher is great”, and “I am fond of my teacher”; Cronbach’s  $\alpha = .92$ ,  $ICC = .15$ ). These items were formulated as simple as possible in order to be understandable for third grade students. We also applied these items in a pilot study to  $N = 159$  students from six classes of grade level two and three. This study revealed a good reliability of the scale, and students did not show any problems with understanding the items. All student survey items were read aloud to the students to avoid language and reading difficulties. Student surveys and tests were applied by trained research assistants. To measure students’ prior interest in science education, we used a four-item scale (e.g., “I put effort into science class because it is fun”; Cronbach’s  $\alpha = .89$ ,  $ICC = .20$ ) that was based on a scale by Blumberg (2008). Student interest after the science classes was measured with a similar scale that was formulated to focus on students’ interest in the teaching unit (e.g., “I put effort into the topic of floating and sinking because it was fun”; Cronbach’s  $\alpha = .91$ ,  $ICC = .16$ ).

### 4.2.2 Teacher questionnaires

Teacher self-reports were assessed at Measurement Point A (Fig. 2). Teacher self-efficacy was measured with an established instrument from Schmitz and Schwarzer (2000; nine items; e.g., “I can keep calm even if my course is disrupted”; Cronbach’s  $\alpha = .83$ ). We measured enthusiasm with a six-item scale by Kunter (2008; e.g., “Teaching is a great pleasure for me”; Cronbach’s  $\alpha = .81$ ). Teachers’

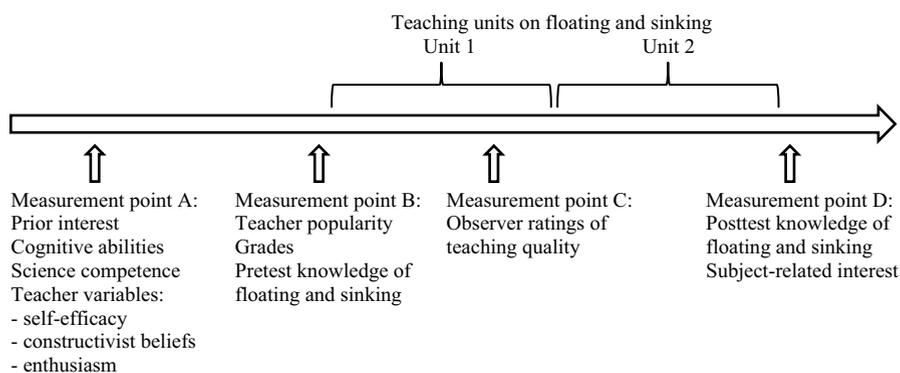


Fig. 2 Study design

constructivist beliefs were measured with a six-item scale that was based on scales by Warwas et al. (2011) and Staub and Stern (2002; e.g., “Children learn especially well when they are allowed to develop their own ideas and go their own way while learning”; Cronbach’s  $\alpha = .63$ ). All of the questionnaire items were rated on 4-point Likert scales ranging from 1 (*strongly disagree*) to 4 (*strongly agree*). One exception was the self-efficacy scale. Here, we kept Schmitz and Schwarzer’s (2000) metric, which ranged from 0 to 100% agreement with the items. The teachers also provided students’ midterm grades. German grades range from 1 (outstanding) to 6 (insufficient). Students received their grades a few days prior to their ratings of teacher popularity.

#### 4.2.3 Standardized tests

We assessed students’ prior science competence with an adapted version of the TIMSS test (Martin et al. 2008) that fit the IPL-Rasch Model (13 items; EAP/PV reliability = .70). Cognitive abilities were assessed with the CFT 20-R (56 items, Cronbach’s  $\alpha = .72$ ; Weiß 2006), a German version of the Culture Fair Intelligence Tests. Students’ conceptual understanding of floating and sinking was assessed with standardized tests. Test items were adapted from existing instruments by Hardy et al. (2006) and Kleickmann et al. (2010). The pretest comprised 16 items (EAP/PV reliability = .52), and the posttest comprised 13 items (EAP/PV reliability = .76). These items have been shown to be sensitive to instruction (Naumann et al. 2017). In addition, experts from educational practice and research in science education have judged the items as valid and highly relevant to the topic of floating and sinking. Items were scored dichotomously or polytomously, and the two tests were scaled separately by applying the Partial Credit Model each time. Student parameters were estimated with weighted likelihood estimates (Warm 1989). All reliabilities reported in this section refer to the dataset of the present study.

#### 4.2.4 Standardized classroom observations

External observers rated teaching quality on three high-inference items: “challenging tasks and questions” (cognitive activation), “recognition and respect towards students” (supportive climate), and “dealing with disruptions and discipline” (classroom management). Previous studies confirmed the validity of these rating items in the prediction of student learning (Fauth et al. 2014b) and students’ ratings of teaching quality (Fauth et al. 2016). Items were rated on a 4-point scale. Raters received extensive training (approximately 40 hr) and assigned their ratings according to a coding manual. Interrater reliability was sufficient ( $ICC > .70$  for two independent raters; Shrout and Fleiss 1979; Wirtz and Caspar 2002).

### 4.3 Design

Figure 2 presents an overview of the research design and measurement points. The analyses for addressing Research Question 1 evaluated the statistical properties

of the teacher popularity scale applied at Measurement Point B. The analyses for addressing Research Questions 2a and 2b used student and teacher data measured at Point A to predict student ratings of teacher popularity at Point B. For Research Question 2c (relations with teaching quality), we examined correlations between teacher popularity and ratings of teaching quality by external observers at Point C. We did not integrate these ratings into the aforementioned regression models because the ratings of teaching quality took place after the student ratings of teacher popularity.

Research Questions 2 and 3 were addressed in a longitudinal design that enabled us to examine students' development during two predesigned teaching units. For Research Question 3, we used student ratings of teacher popularity (Point B) to predict students' learning gains and the development of subject-related interest after the two units (Point D), controlling for prestudy performance variables (Point A). This procedure had the advantage of higher power to detect the effects of teacher popularity as there were fewer external uncontrolled factors that could influence the results compared with examining student development across longer periods. The limitations of our approach are discussed in Sect. 6.5.

The longitudinal study was part of a larger design for evaluating different teaching approaches in science education in German primary schools. In the current study, the teachers taught two predesigned teaching units on floating and sinking, each consisting of nine lessons (45 min each) that were integrated into regular courses for a duration of about nine weeks. The teaching units were adapted from an empirically evaluated science curriculum for teaching floating and sinking. The curriculum was modelled on the principles of inquiry-based science education (Hardy et al. 2006). The first unit covered the concept of density; the second unit focused on the concepts of buoyancy force and displacement.

#### 4.4 Data analyses

We computed the ICC2 index (Bliese 2000; Lüdtke et al. 2009) to examine whether teacher popularity could be reliably assessed at the classroom level (Research Question 1). ICC2 built on ICC1 but also took into account the number of students per class. Thus, the index accounts for the fact that a classroom composite can be measured more reliably when more students provide ratings.

To examine Research Question 2, we computed multilevel regression analyses with students' liking of the teacher as the dependent variable. The individual student characteristics were introduced as grand-mean-centered Level 1 predictors, and the teacher characteristics were Level 2 predictors. Observer ratings of teaching quality were not included in these regressions as teaching quality was assessed after teacher popularity. Thus, we interpreted bivariate correlations for Research Question 2c.

To examine Research Question 3, we computed two different multilevel regression models with students' posttest scores on the achievement test or students' post-interest scores as dependent variables. Individual achievement covariates (pretest, science competence, and cognitive abilities) were introduced as group-mean-centered Level 1 predictors and additionally as grand-mean-centered Level

2 predictors. These Level 2 variables were manifest classroom aggregates of individual variables. Therefore, the covariates accounted for variance within classes as well as variance between classes. With regard to Research Question 3b (different results for different levels of analysis), teacher popularity was introduced at Level 2 as a classroom aggregate. At the individual level, students' liking of the teacher was introduced as a group-mean-centered predictor (Lüdtke et al. 2009). To examine Research Question 3c, we introduced observer ratings of teaching quality to examine the effect of teacher popularity over and above teaching quality.

All regressions were estimated in Mplus 7 (Muthén and Muthén 1998–2012) as doubly manifest models according to Marsh et al. (2012) framework. This approach was expected to provide the most accurate estimates for our data set as we had a relatively small sample size at Level 2, and our sample included the vast majority of students in each class (96%).

The issue of missing values requires careful consideration (Enders 2010). In our study, a relatively small amount of missing data occurred at the level of individual students (average 8.2%, range 6.8–9.7%). Missing data on teacher questionnaires occurred in three cases. Missing values were generated when students or teachers did not attend school on the day the measurements were taken. For one of the 54 classrooms, no observations or video recordings could be made for organizational reasons. There was no indication of a systematic accumulation of missing data patterns across scales or measurement points. No missing data occurred for classroom-level aggregates of individual student data. We used a full information maximum likelihood algorithm (FIML; Arbuckle 1996) to deal with missing data in all regression models.

**Table 1** Correlations and descriptive statistics—individual level

	(1)	(2)	(3)	(4)	(5)	(6)	<i>M</i>	<i>SD</i>
(1) Teacher popularity	–						3.56	0.77
(2) Gender	–.20*	–					–	–
(3) Grades	–.06	.03	–				2.32	0.89
(4) Science competence	–.05	.07*	–.47*	–			–0.32	1.07
(5) Cognitive abilities	.03	–.05	–.42*	.39*	–		104.03	14.77
(6) Pretest	–.06	.14*	–.23*	.24*	.23*	–	0.02	0.90
(7) Posttest	.08*	.08*	–.43*	.38*	.36*	.28*	0.01	0.95

\* $p < .05$

**Table 2** Correlations and descriptive statistics—classroom level

	(1)	(2)	(3)	(4)	(5)	(6)	<i>M</i>	<i>SD</i>
(1) Teacher popularity	–						3.54	0.34
(2) Supportive climate	.41*	–					3.07	0.73
(3) Classroom management	.13	.55*	–				3.37	0.78
(4) Cognitive activation	.04	.28*	.45*	–			3.19	0.82
(5) Constructivist beliefs	.11	.01	–.03	–.02	–		3.24	0.34
(6) Enthusiasm for teaching	.38*	.12	–.07	–.05	.02	–	3.48	0.38
(7) Self-efficacy	.41*	.35*	.10	.30*	.29*	.33*	73.19	9.59

\* $p < .05$

## 5 Results

### 5.1 Descriptive statistics

Tables 1 and 2 show descriptive statistics and correlations between the variables at the individual level and the classroom level, respectively. All of the scales showed relatively high mean scores. Standard deviations for the classroom-aggregated teacher popularity scores were considerably smaller than individual students' reports of how much they liked the teacher. However, these standard deviations at the classroom level were comparable to those obtained from teachers' self-reports (Table 2).

### 5.2 Assessment of teacher popularity at the classroom level

The proportion of variance in students' liking of the teacher that could be attributed to the classroom level was 15%. This is within the range of what is usually observed in students' ratings of teaching. An average of 20 students were assessed per class (ranging from 10 to 27). This resulted in an ICC2 index of .74. This does not reflect perfect agreement between students, but it is above the threshold of .70 that is taken to indicate sufficient agreement (LeBreton and Senter 2008; Lüdtke et al. 2009).

### 5.3 Relations with teacher and student characteristics

Research Questions 2a and 2b asked for the antecedents of teacher popularity in terms of student and teacher characteristics. These were examined in multi-level regression analyses that predicted student ratings of teacher popularity as the dependent variable (Table 3). In Model 1, we introduced only students' individual background variables as predictors. The estimate for student gender was negative (i.e., girls scored higher on the teacher popularity scale). Remarkably, the midterm grades that students had received a few days earlier did not affect the degree to which they liked the teacher.

In Models 2 to 4 (Table 3), teacher characteristics were introduced as predictors in three separate models. Enthusiasm for teaching and self-efficacy were significantly related to teacher popularity ratings after controlling for students' individual

**Table 3** Multilevel regression analyses predicting teacher popularity from student and teacher characteristics

Predictor	Model 1	Model 2	Model 3	Model 4	Model 5
<i>Individual level (students)</i>					
Student gender (1 = male)	-.20 (.03)*	-.20 (.03)*	-.20 (.03)*	-.20 (.03)*	-.20 (.03)*
Grade	-.03 (.03)	-.03 (.03)	-.04 (.03)	-.03 (.03)	-.03 (.03)
<i>Classroom level (teacher)</i>					
Constructivist beliefs	–	.17 (.17)	–	–	.07 (.15)
Enthusiasm for teaching	–	–	.45 (.13)*	–	.35 (.15)*
Self-efficacy	–	–	–	.48 (.16)*	.38 (.18)*
R <sup>2</sup> (within)	.04	.04	.04	.04	.04
R <sup>2</sup> (between)	–	.03	.20	.23	.27

Standardized regression weights; standard errors are in parentheses

\* $p < .05$

characteristics. This was not the case for teachers' constructivist beliefs. In Model 5, all predictors were introduced simultaneously to test for their unique contribution toward predicting the outcome. In this model, enthusiasm for teaching and self-efficacy showed unique relations with teacher popularity.

#### 5.4 Relations with teaching quality

Correlations presented in Table 2 show that only observer ratings of supportive climate were significantly related to teacher popularity (Research Question 2c). Classroom management and cognitive activation were not associated with popularity.

#### 5.5 Prediction of student achievement and interest

Research Questions 3a to 3c were examined with two sets of multilevel regression analyses: one predicting student achievement and one predicting students' subject-related interest. We introduced students' liking of the teacher as a predictor at the individual level (group-mean-centered) and teacher popularity at the classroom level (manifest aggregation; see Sect. 4.4). In Model 1 (predicting achievement), we controlled for pretest scores, science competence, and cognitive abilities (see Table 4). The introduction of teacher popularity revealed significant effects at the classroom level (Research Question 3b). By contrast, analyses on the prediction of student interest (controlling for previous interest, Model 3) revealed significant effects at both levels of analysis (Research Question 3b). Thus, for student achievement, only teacher popularity was relevant, whereas for student interest, students' individual liking of the teacher was relevant as well.

In Models 2 and 4, we also controlled for the observer ratings of teaching quality. Adding these variables to the regression model did not weaken the effects of teacher popularity on achievement and interest. Thus, the effects of teacher popularity on

**Table 4** Multilevel regression analyses predicting student achievement and interest from teacher popularity and observed teaching quality

Predictor	Model 1	Model 2	Model 3	Model 4
Dependent variable	Achievement	Achievement	Interest	Interest
<i>Individual level</i>				
Pretest score	.22 (.03)*	.22 (.03)*	.16 (.04)*	.16 (.04)*
Science competence	.28 (.03)*	.28 (.03)*	–	–
Cognitive abilities	.21 (.03)*	.21 (.03)*	–	–
Teacher popularity	.03 (.03)	.03 (.03)	.17 (.03)*	.17 (.04)*
<i>Classroom level</i>				
Pretest score	.03 (.15)	.05 (.14)	.18 (.17)	.20 (.15)
Science competence	.17 (.14)	.13 (.12)	–	–
Cognitive abilities	.35 (.15)*	.26 (.16)*	–	–
Teacher popularity	.24 (.13)*	.29 (.14)*	.62 (.12)*	.65 (.11)*
Cognitive activation	–	.33 (.13)*	–	–.07 (.11)
Supportive climate	–	–.10 (.13)	–	–.20 (.15)
Classroom management	–	.20 (.17)	–	.09 (.16)
R <sup>2</sup> (within)	.17	.17	.06	.06
R <sup>2</sup> (between)	.21	.33	.42	.51

Standardized regression weights; standard errors are in parentheses

\* $p < .05$ , one-tailed test

student outcomes could not be explained by the quality of teaching (Research Question 3c) (Table 4).

## 6 Discussion

In order to empirically explore the concept of teacher popularity, we related it to a well-established model of teacher competence and teaching quality (see Fig. 1). Results showed that the theoretical connections could also be confirmed empirically. Teacher popularity was empirically connected to several of the constructs in this model in a reasonable way. Our results suggest that it is valuable to consider teacher popularity as a separate construct that differs from teaching quality. Moreover, our results on the prediction of student outcomes supported this idea as teacher popularity provided predictive power for student development that was not inherent to teaching quality. In the following sections, we discuss our results in detail.

## 6.1 Assessment of teacher popularity at the classroom level

The reliable assessment of teacher popularity at the classroom level was a prerequisite for most of the following analyses. If there had been no variability at the classroom level, it would have been senseless to estimate correlations with other classroom-level variables such as teachers' self-reports or classroom observations. The ICC2 index indicates that the extent to which students in the same class liked their teacher showed a sufficient degree of agreement. Indeed, the ICC2 values were comparable to values that researchers obtain when they ask for teaching quality (a "climate" construct according to Marsh et al. 2012). These results led us to the conclusion that there is indeed a *teacher popularity* construct that is assessable at the classroom level (i.e., we can treat teacher popularity as a characteristic of teachers, not merely of individual student–teacher relationships). As one would expect, the teacher popularity scores at the classroom level were very high. However, even in Grade 3, some teachers were more popular with their students than others.

## 6.2 Effects of teacher popularity on students' interest and learning

Our results indicate that teachers differ in their popularity, and these differences are relevant for learning outcomes—teacher popularity matters as a characteristic of teachers (Research Question 3b). A closer look at the different outcomes showed that the effect of teacher popularity on student achievement was limited to the classroom level. By contrast, effects on the development of student interest could be found at both levels of analysis. Thus, there is an additional benefit of an individual positive student–teacher relationship for student interest. A student who—compared with his or her classmates—reports liking the teacher more will also be more interested in the teaching units after pre-existing subject-related interest is controlled for.

These results are in line with previous findings by Montalvo et al. (2007) who reported positive relations between liking the teacher and student outcomes (motivation and semester grades). However, our study extended this knowledge because we focused on elementary schools, whereas previous research was concerned only with high-school teaching. In addition, we were able to examine the effects of *teacher popularity* rather than only the extent to which individual students liked the teacher, and we used standardized measures of achievement rather than only student-reported semester grades.

With regard to student achievement, Wagner (2008)—who examined correlations between popularity and student achievement at both levels of analysis—reported significant associations only at the individual level. The bivariate correlations between popularity and posttest scores reported in Table 1 are in line with these results. However, our study extended this knowledge because, with our longitudinal measurement design, we were also able to confirm an effect of teacher popularity at the classroom level. An important difference between Wagner's (2008) study and ours is the age of the students (third graders vs. ninth graders). We can assume that the teacher–student relationship plays a special role in primary school. Here, the teacher is not only an instructor but also a caregiver and educator. As it seems, popularity is

more important for primary school teachers than it is for secondary school teachers. However, we certainly need more research on the issue of different grade levels and teacher popularity.

### 6.3 Unique contribution over and above teaching quality

What are the mechanisms behind the effects on student outcomes? As pointed out earlier, classroom processes may determine student ratings of teacher popularity to a large extent. The classroom is the place where the students and teacher meet and where they interact. It is thus plausible that teachers who are better at teaching are also more popular with the students. However, previous research had not determined whether the positive effects of teacher popularity could be attributed merely to the fact that popular teachers are better at teaching. The results of our study show that this is not the case. There is something about popular teachers that is not captured by teaching quality ratings but is nonetheless relevant for student learning. Results from a previous study showed that this also held true when student ratings of teaching quality were controlled for (Fauth et al. 2014a). With regard to the scientific discussion on halo effects, we can assume that global ratings of teacher popularity contain “valid halo” variance (Lance and Woehr 1986) that is not captured by specific measures of classroom quality.

Substantively, our results suggest that popular teachers are able to motivate their students and awaken students’ interest in the subject matter. The findings on the positive effect of teacher popularity on the development of student interest support this idea. With regard to expectancy-value theories (Wigfield and Eccles 2000), we can assume that popular teachers have a positive impact on the value component of motivation. A second factor that has the potential to influence student motivation beyond the quality of teaching is the mechanism of “pleasing the teacher” (Montalvo and Roedel 1995). Popular teachers are probably also the ones students do not want to disappoint with poor performances, and such a desire could lead to greater effort and, in turn, to better learning results. Montalvo et al.’s (2007) results support this hypothesis. In their study, “effort” and “persistence” were among the scales with the largest differences between liked and disliked teachers.

### 6.4 Associations with teacher characteristics, classroom observations, and students’ background

The analyses for addressing Research Question 2 were concerned with the conditions that may have produced differences in teacher popularity at the classroom level and differences in the extent to which individual students liked the teacher.

The relations between teacher popularity and observed teaching quality revealed that—as expected—only supportive climate was significantly associated with teacher popularity. External observers’ ratings of supportive climate cover aspects of student–teacher interactions such as respect, warmth, and the recognition of students. Its focus is on social and emotional aspects of student–teacher interactions in

the classroom, whereas the other dimensions aim at students' in-depth understanding of subject-specific concepts (cognitive activation) or at ensuring order and structure in the class (classroom management).

It makes sense that supportive climate and teacher popularity would be related to each other. A supportive teacher's behavior in the classroom may lead to popularity, or it might be easier for popular teachers to establish a supportive climate (thus, there are arrows in both directions in Fig. 1). But the two are not interchangeable as one refers to the quality of teaching, and the other refers to the class' affectively tinged general impression of their teacher.

The multilevel regression analyses revealed that teacher popularity was associated with teachers' enthusiasm and self-efficacy but not with their constructivist beliefs. It makes sense that teachers who report feeling enthusiastic about teaching are more popular with students. Empirical studies show that enthusiasm for teaching is also related to teaching quality: Enthusiastic teachers provide more learning support and are better able to manage the classroom (Kunter et al. 2008). It is possible that the connection between enthusiasm and teacher popularity is mediated by teacher behaviors such as individual learning support. Our result on the empirical connection between teacher popularity and the observed supportive climate in the classroom is in line with this idea. We can assume similar mechanisms for teacher self-efficacy (Skaalvik and Skaalvik 2007). In previous studies, teacher self-efficacy was empirically related to teaching quality (Guo et al. 2014; Holzberger et al. 2013; Justice et al. 2008).

Highly motivated and self-confident teachers probably have a more expressive teaching style, an idea that links the present results to the discussion about the Dr. Fox effect (Marsh 1987). However, after the discussion of results on the prediction of student outcomes, we argue—in line with modern interpretations of the Dr. Fox effect—that such an effect does not necessarily lead students to provide useless teacher ratings.

In contrast to the findings for enthusiasm and self-efficacy, teachers' constructivist beliefs did not significantly predict popularity. The results of previous studies on teachers' constructivist beliefs have been somewhat mixed (Dubberke et al. 2008; Kunter et al. 2013). This might be due to the fact that certain beliefs do not always lead to a corresponding teaching practice. Another plausible interpretation might be that students are actually able to identify differences in teaching practices, but such differences might not be relevant to the extent to which a student likes the teacher. Constructivist learning settings might be more exciting but also more demanding and perhaps more exhausting, which would lead to nonsignificant relations between constructivist beliefs and teacher popularity.

The above-mentioned teacher variables can explain differences in teacher popularity only between classes. The individual student background variables can also explain differences within classes (i.e., the extent to which each individual student likes the teacher). The results revealed that girls reported liking their teachers more than boys did. This result is in line with previous findings that girls rate teachers' behaviors in a more positive manner (Centra and Gaubatz 2000; Wagner 2008). However, we cannot determine whether this effect is specific to girls' ratings of female teachers or teachers in general as the great

majority (86%) of teachers in our sample were female (corresponding to the gender composition in German elementary schools). In addition, previous studies have also found student-gender by teacher-gender interactions (Centra and Gaubatz 2000) such that girls tend to prefer female teachers, another trend that might have contributed to our results.

As pointed out in the theory section, grading leniency has been discussed as a potential bias in students' ratings of teaching quality (Benton and Cashin 2012; Marsh 1987). It is interesting to note that the grades that students received were not related to how much they liked the teacher. We can think of two plausible explanations for this result: The teachers' grading practices might indeed have been fair and in accordance with students' actual performance, or students in Grade 3 are not able to detect unfair grading practices. In any case, one important message of the present investigation is that teachers' popularity in elementary school does not depend on their individual grading practices.

## 6.5 Overall strengths and limitations

The present study adds knowledge to an important field of research by investigating the phenomenon of teacher popularity in primary schools. Although students' liking of a teacher and teacher popularity play a particularly important role in the earlier stages of students' academic development, previous research has primarily concentrated on secondary schools (e.g., Montalvo et al. 2007). We could show that some primary school classes rated their teacher to be more popular than other classes. Although popularity ratings were generally high, primary school teachers were not equally popular—and these differences in popularity were associated with differences in students' academic development.

A major strength of the present investigation is the variety of sources of data we were able to apply (see Fig. 1). Instead of relying only on student ratings, which might have led to flawed correlations between the variables of interest, we were able to apply the most appropriate measurement procedures for each construct: teacher self-reports for teacher variables, video-observer ratings for teaching quality, and student ratings for teacher popularity. Another important feature of the study goes along with another major advantage and a drawback as well. We drew on a highly standardized design: The development of student achievement and motivation were assessed during two teaching units (of in total 9 weeks duration). The teaching units were predesigned with regard to sequencing and materials. As we also controlled for student performance before the units began, variability in student outcomes after the units could more easily be attributed to teacher popularity. However, this goes along with a major limitation of the study. The focus on development during only two teaching units allowed for the investigation of only short-term effects.

Another limitation is the cross-sectional nature of our analyses on the antecedents of teacher popularity. The relations with teacher and teaching characteristics were only correlational, and thus, causal interpretations are not warranted. The study cannot provide any evidence on the development of teacher popularity.

For example, it is plausible that teacher enthusiasm causally influences popularity, but our data did not allow us to really test this hypothesis. Accordingly, the opposite direction (teacher popularity influences enthusiasm) would be equally plausible. Future studies should take into account both causal directions and their development over time.

Future studies might also be able to examine the important question of the stability of teacher popularity between different classes taught by the same teacher. Our study included one teacher per class, and it was only this class that rated the teacher's popularity. Our test of agreement between students thus included only students from the same class. Research has yet to determine whether there must be a certain fit between teacher and class in order for popularity to emerge. Additionally, it is still an open question whether teacher popularity is a stable personal characteristic of a teacher or if it changes over time in professional developments. As popularity is related to teachers' professional competence and teaching behavior, changes in these areas (cf. Malmberg et al. 2010) might also go along with changes in popularity.

## 6.6 Conclusion

Popular teachers are highly motivated to teach, and they show self-confidence in class even in stressful situations. It seems that their professional subject-related beliefs are less important for popularity—at least in primary school. Popular teachers are able to create a supportive climate in the classroom, treat students with respect, and care about their students' problems. Other dimensions of teaching quality (i.e., classroom management and cognitive activation) are less relevant for teacher popularity. The information obtained from student ratings of teacher popularity is related to student learning and motivation over and above the effects of teaching quality in the classroom. Knowing the popularity of a teacher gives us relevant information that is not inherent to measures of teaching quality, and this means that teacher popularity is a useful indicator of teacher quality, and it should be considered in future research.

**Acknowledgments** This research was funded by the Hessian initiative for the development of scientific and economic excellence (LOEWE).

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Aleamoni, L. M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education*, *13*, 153–166.
- Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 243–277). Mahwah: Lawrence Erlbaum.
- Atamian, R., & Ganguli, G. (1993). Teacher popularity and teaching effectiveness: Viewpoint of accounting students. *Journal of Education for Business*, *68*(3), 163–170.
- Bandura, A. (1995). *Self-efficacy in changing societies*. Cambridge: Cambridge University Press.
- 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.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., et al. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, *47*(1), 133–180.
- Benton, S. L., & Cashin, W. E. (2012). *Student ratings of teaching: A summary of research and literature. IDEA Paper #50*. Manhattan: The IDEA Center.
- Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. Kozlowski (Eds.), *Multilevel theory, research, and methods in organisations* (pp. 149–381). San Francisco: Jossey-Bass.
- Blumberg, E. (2008). *Multikriteriale Zielerreichung im naturwissenschaftsbezogenen Sachunterricht der Grundschule* (Multi-criterial goal attainment in science education in primary school). Doctoral dissertation. University of Münster, Germany.
- Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education*, *44*, 495–518.
- Centra, J. A., & Gaubatz, N. B. (2000). Is there a gender bias in student evaluations of teaching? *Journal of Higher Education*, *70*, 17–33.
- Corey, S. M., & Beery, G. S. (1938). The effect of teacher popularity upon attitude toward school subjects. *The Journal of Educational Psychology*, *29*(9), 665–670.
- Cornelius-White, J. (2007). Learner-centered teacher-student relationships are effective: A meta-analysis. *Review of Educational Research*, *77*(1), 113–143.
- Davis, H. A. (2003). Conceptualizing the role and influence of student-teacher relationships on children's social and cognitive development. *Educational Psychologist*, *38*(4), 207–234.
- Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: human needs and the self-determination of behavior. *Psychological Inquiry*, *11*(4), 227–268.
- Decristan, J., Hondrich, A. L., Büttner, G., Hertel, S., Klieme, E., Kunter, M., et al. (2015). Impact of additional guidance in science education on primary students' conceptual understanding. *The Journal of Educational Research*, *108*, 358–370.
- Dubberke, T., Kunter, M., McElvany, N., Brunner, M., & Baumert, J. (2008). Lerntheoretische Überzeugungen von Mathematiklehrkräften: Einflüsse auf die Unterrichtsgestaltung und den Lernerfolg von Schülerinnen und Schülern (Mathematics teachers' beliefs and their impact on instructional quality and student achievement). *Zeitschrift für Pädagogische Psychologie*, *22*(3), 193–206.
- Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., et al. (1993). Development during adolescence. The impact of stage-environment fit on young adolescents' experiences in schools and in families. *American Psychologist*, *48*, 90–101.
- Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014a). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, *29*, 1–9.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014b). Grundschulunterricht aus Schüler-, Lehrer- und Beobachterperspektive (Teaching quality in primary school from the perspectives of students, teachers, and external observers). *Zeitschrift für Pädagogische Psychologie*, *28*(3), 127–137.

- Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2016). Urteile zur Unterrichtsqualität von Grundschüler/innen: Mehr als, Ich mag meine Lehrerin“? [Student ratings of teaching quality: More than just “I like my teacher”?]. In K. Liebers, B. Landwehr, S. Reinhold, S. Riegler, & R. Schmidt (Eds.), *Facetten grundschulpädagogischer und grundschuldidaktischer Forschung* (S. 157–162). Wiesbaden: Springer.
- Feldman, K. A. (1986). The perceived instructional effectiveness of college teachers as related to their personality and attitudinal characteristics: A review and synthesis. *Research in Higher Education*, 24(2), 139–213.
- Gruehn, S. (2000). *Unterricht und schulisches Lernen. Schüler als Quellen der Unterrichtsbeschreibung [Teaching and learning in school. Students as data sources]*. Münster: Waxmann.
- Guo, Y., Dymia, J. M., Yeager Pelatti, C., & Justice, L. M. (2014). Self-efficacy of early childhood special education teachers: Links to classroom quality and children’s learning for children with language impairment. *Teaching and Teacher Education*, 39, 12–21.
- Hardy, L., Jonen, A., Möller, K., & Stern, E. (2006). Effects of instructional support within constructivist learning environments for elementary school students’ understanding of “floating and sinking”. *Journal of Educational Psychology*, 98, 307–326.
- Hardy, L., Hertel, S., Kunter, M., Klieme, E., Warwas, J., Büttner, G., et al. (2011). Adaptive Lerngelegenheiten in der Grundschule: Merkmale, methodisch-didaktische Schwerpunktsetzungen und erforderliche Lehrkompetenzen [Adaptive learning environments in primary school]. *Zeitschrift für Pädagogik*, 57, 819–833.
- Holzberger, D., Philipp, A., & Kunter, M. (2013). How teachers’ self-efficacy is related to instructional quality: A longitudinal analysis. *Journal of Educational Psychology*, 105(3), 774–786.
- Justice, L. M., Mashburn, A. J., Hamre, B. K., & Pianta, R. C. (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early Childhood Quarterly*, 23, 51–68.
- Kleickmann, T., Vehmeyer, J., & Möller, K. (2010). Zusammenhänge zwischen Lehrervorstellungen und kognitivem Strukturieren im Unterricht [Relations between teacher conceptions and features of scaffolding]. *Unterrichtswissenschaft*, 38(3), 210–228.
- Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janik & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137–160). Münster: Waxmann.
- Kunter, M., Frenzel, A., Nagy, G., Baumert, J., & Pekrun, R. (2011). Teacher enthusiasm: Dimensionality and context specificity. *Contemporary Educational Psychology*, 36, 289–301.
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teacher: Effects on instructional quality and student development. *Journal of Educational Psychology*, 105(3), 805–820.
- Kunter, M., Tsai, Y.-M., Klusmann, U., Brunner, M., Krauss, S., & Baumert, J. (2008). Students’ and mathematics teachers’ perceptions of teacher enthusiasm and instruction. *Learning and Instruction*, 18(5), 468–482.
- Kyriakides, L., Christoforou, C., & Charalambous, C. Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education*, 36, 143–152.
- Lance, C. E., & Woehr, D. J. (1986). Statistical control of halo: Clarification from two cognitive models of the performance appraisal process. *Journal of Applied Psychology*, 71, 679–685.
- LeBreton, J. M., & Senter, J. L. (2008). Answers to twenty questions about interrater reliability and interrater agreement. *Organizational Research Methods*, 11, 815–852.
- Lee, J.-S. (2012). The effects of the teacher–student relationship and academic press on student engagement and academic performance. *International Journal of Educational Research*, 53, 330–340.
- Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students’ understanding of the Pythagorean theorem. *Learning and Instruction*, 19(6), 527–537.
- Lütke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary Educational Psychology*, 34(2), 120–131.
- Lui, A. M., & Bonner, S. M. (2016). Preservice and inservice teachers’ knowledge, beliefs, and instructional planning in primary school mathematics. *Teaching and Teacher Education*, 56, 1–13.

- Malmberg, L.-E., Hagger, H., Burn, K., Mutton, T., & Colls, H. (2010). Observed classroom quality during teacher education and two years of professional practice. *Journal of Educational Psychology, 102*(4), 916–932.
- Marsh, H. W. (1987). Students' evaluation of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research, 11*, 253–388.
- Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J. S., Abduljabbar, A. S., et al. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist, 47*, 106–124.
- Marsh, H. W., & Ware, J. E. (1982). Effects of expressiveness, content coverage, and incentive on multidimensional student rating scales: New interpretations of the Dr. Fox effect. *Journal of Educational Psychology, 74*, 126–134.
- Martin, M. O., Mullis, I. V. S., & Foy, P. (2008). *TIMSS 2007 international science report*. Chestnut Hill: TIMSS & PIRLS International Study Center.
- Montalvo, G. P., Mansfield, E. A., & Miller, R. B. (2007). Liking or disliking the teacher: Student motivation, engagement and achievement. *Evaluation and Research in Education, 20*(7), 144–158.
- Montalvo, G. P. & Roedel, T. (1995). *Pleasing the teacher: A qualitative look*. Paper presented at the annual meeting of AERA, San Francisco, CA.
- Muthén, L. K. & Muthén, B. O. (1998–2012). *Mplus user's guide. Seventh edition*. Los Angeles: Muthén & Muthén.
- Naumann, A., Hartig, J., & Hochweber, J. (2017). Absolute and relative measures of instructional sensitivity. *Journal of Educational and Behavioral Statistics, 42*(6), 678–705.
- Payne, M. A. (1987). Determinants of teacher popularity and unpopularity: A West Indian perspective. *Journal of Education for Teaching, 13*(3), 193–205.
- Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*(2), 109–119.
- Schmitz, G. S., & Schwarzer, R. (2000). Selbstwirksamkeitserwartung von Lehrern: Längsschnittbefunde mit einem neuen Instrument (Perceived self-efficacy of teachers: Longitudinal findings with a new instrument). *Zeitschrift für Pädagogische Psychologie, 14*(1), 12–25.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428.
- Skaalvik, E. M., & Skaalvik, S. (2007). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Journal of Educational Psychology, 99*(3), 611–625.
- Staub, F. C., & Stern, E. (2002). The nature of teachers' pedagogical content beliefs matters for students' achievement gains: Quasi-experimental evidence from elementary mathematics. *Journal of Educational Psychology, 94*, 344–355.
- Wagner, W. (2008). *Methodenprobleme bei der Analyse von Unterrichtswahrnehmung aus Schülersicht* (Methodological issues in analysing students' perceptions of teaching). Doctoral dissertation. Germany: University of Koblenz-Landau. Retrieved from <http://d-nb.info/987591800/34>. Accessed 31 Oct 2017.
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54*, 427–450.
- Warwas, J., Hertel, S., & Labuhn, A. S. (2011). Bedingungsfaktoren des Einsatzes von adaptiven Unterrichtsformen im Grundschulunterricht (Conditions for the application of adaptive teaching in elementary school). *Zeitschrift für Pädagogik, 57*, 854–867.
- Weiß, R. H. (2006). *CFT 20-R. Grundintelligenztest Skala 2—Revision [Culture fair test]*. Göttingen: Hogrefe.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology, 25*, 68–81.
- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität [Interrater agreement and interrater reliability]*. Göttingen: Hogrefe.
- Zee, M., Koomen, H., Jellesma, F., Geerlings, J., & de Jong, P. (2016). Inter- and intra-individual differences in teachers' self-efficacy: A multilevel factor exploration. *Journal of School Psychology, 55*, 39–56.

**Benjamin Fauth** is assistant professor at the Hector Research Institute of Education Sciences and

Psychology at the University of Tübingen, Germany, and a member of IDeA in Frankfurt, Germany. His research interests include the assessment of teaching quality in schools and the effects of teacher competence and teaching quality on student development.

**Jasmin Decristan** is professor at the School of Education of the University of Wuppertal and a member of IDeA in Frankfurt, Germany. Her research focuses on individual support of students, differential effects of instruction, teaching quality, and peer-assisted learning.

**Svenja Rieser** is research assistant at the School of Education of the University of Wuppertal. Her research interests are teaching quality, self- and shared regulation of learning, and early education.

**Eckhard Klieme** is director of the Department of Educational Quality and Evaluation at the DIPF, a member of the scientific board of IDeA in Frankfurt, Germany, and professor of educational science at the Goethe University in Frankfurt. His research focuses on teaching quality, school effectiveness, school development, and international comparative education research.

**Gerhard Büttner** is professor of educational psychology at the Goethe University, a member of IDeA in Frankfurt, Germany, and head of the Competence Center School Psychology Hesse. His research focuses on self-regulated learning, teaching quality, working memory, learning disabilities, and intellectual disabilities.

## Affiliations

**Benjamin Fauth**<sup>1,2</sup>  · **Jasmin Decristan**<sup>2,3</sup> · **Svenja Rieser**<sup>3</sup> · **Eckhard Klieme**<sup>2,4</sup> · **Gerhard Büttner**<sup>2,5</sup>

Jasmin Decristan  
decristan@uni-wuppertal.de

Svenja Rieser  
rieser@uni-wuppertal.de

Eckhard Klieme  
klieme@dipf.de

Gerhard Büttner  
buettner@paed.psych.uni-frankfurt.de

<sup>1</sup> Hector Research Institute of Education Sciences and Psychology, University of Tübingen, Europastraße 6, 72072 Tübingen, Germany

<sup>2</sup> Center for Individual Development and Adaptive Education of Children at Risk (IDeA), Frankfurt, Germany

<sup>3</sup> University of Wuppertal, Gaußstr. 20, 42119 Wuppertal, Germany

<sup>4</sup> German Institute for International Educational Research Frankfurt, Schloßstr. 29, 60486 Frankfurt, Germany

<sup>5</sup> Goethe University Frankfurt, Theodor-W.-Adorno-Platz 6, 60629 Frankfurt, Germany