Judgment Confidence and Judgment Accuracy of Teachers in Judging Self-Concepts of Students

ANNA-KATHARINA PRAETORIUS
VALÉRIE-DANIELLE BERNER
University of Augsburg, Germany

HORST ZEINZ
University of Muenster, Germany

ANNETTE SCHEUNPFLUG
Friedrich-Alexander-University
Erlangen-Nuremberg, Germany

MARKUS DRESEL
University of Augsburg, Germany

ABSTRACT. Accurate teacher judgments of student characteristics are considered to be important prerequisites for adaptive instruction. A theoretically important condition for putting these judgments into operation is judgment confidence. Using a German sample of 96 teachers and 1,388 students, the authors examined how confident teachers are in their judgments of students' mathematic and verbal self-concepts, and whether judgment confidence is related to judgment accuracy. Judgment confidence was largely student specific, and the majority of teachers were overconfident of their judgments. Moreover, teacher confidence was higher for extreme judgments. In the subject of mathematics, judgment confidence was moderately associated with judgment accuracy. The findings challenge the efficacy of adapting instruction to student characteristics, as it is obvious that many teachers are not aware of their judgment inaccuracy.

Keywords: diagnostic competence, judgment accuracy, judgment confidence, self-concept, teacher

Conducting diagnoses is an important task which is performed by teachers on a daily basis. Planning lessons, organizing student-learning groups, choosing learning materials and so on: All of these instructional activities call for decisions based on assessments of student learning prerequisites to secure proper adaptation to student needs in subsequent instruction (Seidel & Shavelson, 2007). To enable effective instruction, it is important that these assessments are quite accurate. When inaccurate judgments are made of students' abilities, prior knowledge, learning and achievement motivation, and other student characteristics, many instructional activities should become less effective in terms of learning and achievement gains (Anders, Kunter, Brunner, Krauss, & Baumert, 2010). Over and beyond this fundamental assumption, teachers must furthermore be convinced that their judgments are correct to ensure the effective utilization of assessments and judgments. For several decades, researchers in the field of so-

cial psychology have been investigating how confident persons are in the judgments they pass of various characteristics in various contexts (for an overview, see Baker, 2010). However, no investigations concerning the confidence teachers have in the judgments they make of their students' characteristics have been conducted thus far. Nevertheless, evidence regarding teacher assumptions on whether their own judgments of students are correct or not would be highly relevant: According to the aforementioned assumption, judgment confidence among teachers should be significant for the effectiveness of instruction.

To shed light on teacher judgment confidence and its associations with judgment accuracy, we conducted an empirical study in German secondary schools and investigated teachers who judged the mathematics or verbal self-concepts of their students. In the subsequent sections of the introduction, theoretical considerations and empirical findings concerning the relevance of teacher judgment accuracy are summarized; afterward, student self-concepts are brought into focus. Next, the importance of judgment confidence is described and empirical findings concerning the relationship between judgment confidence and judgment accuracy as well as some characteristics of judgment confidence are reported.

Importance of Accurate Teacher Judgments and the Actual (In)Accuracy of Teacher Judgments

There is a consensus in literature concerning the theoretically assumed positive effects of high judgment accuracy among teachers. In accordance with the principle of adaptation to differences between learners (see Corno & Snow, 1986), the basic tenet is that teachers who are versed in

Address correspondence to Anna-Katharina Praetorius, Department of Psychology, University of Augsburg, Universitaetsstraße 10, Augsburg, 86159, Germany. (E-mail: anna.praetorius@phil.uni-augsburg.de)

judging student characteristics are able to adequately and effectively adapt their classroom activities to the learning requirements of their students (e.g., Begeny, Eckert, Montarello, & Storie, 2008; Edelenbos & Kubanek-German, 2004; Hoge & Coladarci, 1989). Preliminary empirical evidence exists to support the theoretically deduced positive effects of judgment accuracy on instructional effectiveness (Helmke & Schrader, 1987; cf. Anders et al., 2010).

Earlier studies have demonstrated that teachers exhibit considerably different levels of accuracy when judging different types of characteristics among their students: Scholastic achievement is one of the characteristics that teachers, on average, can assess relatively well—in the meta-analysis conducted by Hoge and Colardarci (1989), correlations (r) between teacher judgments and actual student achievement ranged between .28 and .92 with a median of .66. According to a study on student test-anxiety conducted by Boehnke, Silbereisen, Reynolds, and Richmond (1986), teacher judgments and actual student characteristics were rather poorly correlated ($r = .21$; no details pertaining to range). The academic self-concepts of students were, on average, judged with moderate accuracy; for example, Spinath (2005) reported correlations (r) between $-.39$ and .82 with a median of .39. Thus, particularly for judgments in areas other than achievement, there is often a relatively low level of accuracy for teacher judgments.

Why Should Teachers Be Able to Judge Students' Self-Concepts Accurately?

The accuracy of judgments of student achievement is of high importance (e.g., with respect to a fair achievement evaluation and fair access to institutes of higher learning). The importance of accurate teacher judgments for other student characteristics may not be quite as obvious. However, student learning prerequisites are also important in planning and conducting high quality instruction. In this regard, academic self-concept is considered to be one of the most important motivational learning prerequisites for students (cf. Möller, Pohlmann, Köller, & Marsh, 2009). The focus of the present work, therefore, is on these self-concepts. A student's academic self-concept is defined as the student's self-perceptions of his or her academic abilities. These self-perceptions can be realistic (i.e., in agreement with the actual abilities), but this is not always the case. According to the hierarchical self-concept model of Shavelson, Hubner, and Stanton (1976), the academic self-concept can be subdivided into domain-specific self-concept components for specific school subjects (e.g., mathematics self-concept, verbal self-concept). However, verbal and mathematic self-concepts are frequently uncorrelated (e.g., Marsh, 1986)[1] and therefore are often investigated separately instead of combining them to form a general academic self-concept.

Fostering adequate to slightly optimistic self-concepts is seen as an important task for teachers, as self-concepts influence diverse outcomes in the learning process such as the quality and quantity of learning activities, learning outcomes, and other components of the students' motivational sets (for an overview, see Marsh, Craven, & McInerney, 2003). Additionally, a favorable self-concept is considered to be valuable in itself, because it is linked to general self-esteem and well-being. Students with low assumptions regarding their abilities should therefore be fostered to develop more positive views of themselves. To enable such support, teachers need information concerning the (domain-specific) self-concepts of their students. Beyond enhancing their self-concepts, it is beneficial for all students when their teachers can take academic self-concepts into account. For example, a mathematics teacher should be providing a student who shows a rather highly pronounced mathematics self-concept with trickier, more complicated exercises that are slightly above the student's skill level. This would contribute to optimal performance development in the student. For a student with the same skill level, but who has little confidence in his or her mathematics abilities, a different approach is called for. Such a student first needs to experience success that can be attributed to his or her abilities. This could be accomplished by working through exercises that are commensurate with, or even slightly below, the present skill level. As academic self-concept and achievement reciprocally influence one another (see Marsh & Craven, 2006), using such instructional approaches should result in positive effects on the development of achievement and learning behavior. Such strategies, individually tailored to the particular student, can be useful and effective—but if, and only if, teachers are able to accurately judge which of their students have favorable self-concepts and which of them have unfavorable self-concepts. This is a nontrivial mission because self-concepts are not directly observable and are rather difficult to judge as they often differ from the actual student performances.

Teacher Judgments and Their Behavior-Related Relevance: Necessity of Judgment Confidence

If we assume that judgment accuracy is a necessary precondition for effective adaptation, it follows that teachers with high judgment accuracy should be able to successfully implement their instructional abilities in the classroom. Conversely, we must then assume that teachers with low judgment accuracy are probably ill-equipped to properly adapt their classroom activities, due to inadequate diagnoses of the individual learning needs of their students. Consequently, this should result in ineffective instruction. These assumptions regarding the interplay of judgment accuracy and adaptation implicitly assume that teachers are, regardless of their true judgment accuracy, confident in the judgments they make and, subsequently, adjust their actions to be in accord with the judgments they pass, which in many cases are unrealistic. If we follow this line of thinking, some teachers may not be confident in the judgments they make, and therefore do not base their decisions pertaining to classroom activities

on their own judgments of student characteristics (for this argumentation, see also Patterson, Foster, & Bellmer, 2001).

In considering all of the possible combinations of judgment confidence and judgment accuracy (Table 1), especially the combination of low judgment accuracy with high judgment confidence can be assumed to have negative effects on instructional effectiveness due to the inadequate adaption of classroom activities to the learning requirements of the individual students. What is the present state of empirical evidence concerning the relationship between judgment accuracy and judgment confidence? Because no empirical findings have yet been published regarding teacher judgment accuracy in combination with judgment confidence, in the following we review related research findings obtained in the field of social psychology. Based on this literature review, we derive assumptions regarding the accuracy of judgments made by teachers in assessing the verbal and mathematics self-concepts of their students.

Relationships Between Judgment Accuracy and Judgment Confidence

Several studies in the field of social psychology have suggested that individuals are more confident in the judgments they make than their actual judgment performance indicates (for overviews, see Lichtenstein, Fischhoff, & Philipps, 1982; Metcalfe, 1998). This phenomenon, which is known as overconfidence bias, is particularly evident in conjunction with judging tasks considered to be moderately to very difficult (DePaulo, Charlton, Cooper, Lindsay, & Muhlenbruck, 1997). As stated previously, it can be assumed that judging the self-concepts of others is relatively complex. Therefore, it seems plausible that the majority of teachers tend to overestimate their own accuracy when assessing the self-concepts of their students.

The relationship between judgment accuracy and judgment confidence has been investigated on two levels. On the one hand, analyses have been conducted on a between-judge level (e.g., Ames, Kammrath, Suppes, & Bolger, 2010; Smith, Archer, & Costanzo, 1991). Because the focus here is on differences between the persons who are judging, the average of all judgments made by a specific judge is related to the mean level of confidence for these judgments. In the present case, average teacher ratings of the self-concepts of all of their students would be interrelated to the average or overall confidence teachers have in their judgments. On the other hand, several studies have investigated the relationship between judgment confidence and judgment accuracy on a within-judge level (e.g., DePaulo et al., 1997; Patterson et al., 2001). In these studies the accuracy of separate judgments of separate targets was correlated with the judges' confidence with respect to their separate judgments. In the present case, the self-concept judgments made by teachers for individual students would be interrelated to their specific confidence in the individual ratings.

On a between-judge level, empirical findings are inconsistent: For example, a study in the field of zero acquaintance research shows that the mean judgment accuracy of persons regarding the personality characteristics of strangers is not significantly correlated with their judgment confidence (Ames et al., 2010). Moreover, an investigation in the field of interpersonal perception conducted by Patterson et al. (2001) was not able to confirm a significant relationship between mean judgment confidence and judgment accuracy. However, a significant, although rather small correlation between these two variables could be detected in a study by Smith et al. (1991) in the same research field ($r = .08$). Vallone, Griffin, Lind, and Ross (1990) were also able to detect a correlation between judgment confidence and judgment accuracy in a series of studies on an aggregate level (between $r_{pb} = .18$ and $r_{pb} = .24$, depending on the substudy). Previous research has thus far indicated modest, if any, relationships between judgment confidence and judgment accuracy on a between-judge level. This leads to the question of whether (slight) relationships could exist between mean judgment confidence and judgment accuracy among teachers.

In an investigation in the area of interpersonal perception on the within-judge level, Smith et al. (1991) found a mean correlation (r) of .26 between judgment accuracy and judgment confidence. Also, in the study by Patterson et al. (2001) referenced previously, which was conducted in the same field, a significant correlation (r) of .13 was found between judgment confidence and judgment accuracy on a within-judge level. However, a meta-analysis conducted by

TABLE 1. Possible Combinations of Judgment Accuracy and Judgment Confidence and Their Theoretically Assumed Effects on Instructional Effectiveness

Judgment confidence	Judgment accuracy	
	Low	High
Low	Veridical low confidence Rather no effect on instructional effectiveness	Underconfidence Rather no effect on instructional effectiveness
High	Overconfidence Potentially negative effects on instructional effectiveness	Veridical high confidence Potentially positive effects on instructional effectiveness

DePaulo et al. (1997) came to the conclusion that judgment confidence and judgment accuracy are not significantly correlated on a within-judge level with regard to the identification of deception. Nevertheless, the majority of the studies conducted in the field of social psychology, using diverse targets, point toward a positive relationship between judgment accuracy and judgment confidence on a within-judge level. On this basis we expected that higher teacher confidence in their judgments would be associated with better judgment accuracy.

Confidence Judgments in the Spotlight: Interindividual and Intraindividual Differences

To gain a more detailed understanding of how confidence judgments function and develop, it is worthwhile to not only delve into their relationships with judgment accuracy, but also to take a closer look at interindividual and intraindividual differences in judgment confidences.

First, it can be assumed that teachers differ according to how confident they are in the judgments they make. In the literature there are several clear indications of interindividual differences (see Klayman, Soll, Juslin, & Winman, 2006; West & Stanovich, 1996). As illustrated by research on rater bias (for summary, see Myford & Wolfe, 2003), persons who are unsure of their own judgments avoid choosing the extremities on an answer scale. If this also applies to teachers judging student characteristics, one should find that teachers who are unsure of their judgments will tend to make midrange selections when judging their students' characteristics, and will therefore overlook cases in which students clearly underestimate or overestimate their abilities.

If using the Realistic Accuracy Model developed by Funder (1999), it can be assumed that teacher judgments can also be intraindividually differentiated with regard to their confidence levels. In Funder's model, four different groups of moderators are specified. The moderator group relevant in this context was labeled *good target* by Funder, and explains the perception among judges that the persons on whom they pass judgments are not equivalently easy to assess. It turned out that, among other things, the perceived magnitude of the characteristic to be assessed plays a significant role in this process: Judges are more confident in high and low judgments than in midrange judgments. Such a curvilinear relationship between judgments and confidence was found, for example, by Fox, Bizman, Hoffman, and Oren (1995). On this basis, it is presumed that teachers, when forming judgments, are more confident when they judge the characteristics of their students to be high or low than they are when they rate these characteristics in the middle of the range.

Research Questions and Hypotheses

The present research addresses the questions of whether, and to what degree, teachers are aware of the accuracy of their judgments of student characteristics. This ques-

tion was pursued with regard to the academic self-concepts held by students. As self-concepts are domain-specific, two school subjects were investigated (mathematics and German[2]). Therefore, the targets of the teachers' judgments were their students' mathematics self-concepts and verbal self-concepts. Additionally, the choice of two domains allows for an examination of the degree to which the findings can be generalized across subjects. Another aim of the present study was to undertake a detailed examination of associations between teacher judgment confidence and judgment accuracy on a within-judge level (i.e., covariations within teachers regarding student- or target-specific judgments) and on a between-judge level (i.e., covariations between average judgments made by teachers and judge-specific confidence). All of the studies we referenced pertaining to judgment confidence had been conducted in areas outside of educational contexts. To the best of our knowledge, the present study is the first application of the judgment confidence concept to the issue of the accuracy of teacher judgments of student characteristics. Specifically, as developed in the previous sections, we are pursuing the following hypotheses:

Hypothesis 1 (H_1): There is substantial intraindividual variation regarding teachers' confidence in their judgments of the mathematics and verbal self-concepts of different students.

H_{2a}: On a within-judge level, mathematics and verbal self-concept judgments in the extreme ranges are associated with high judgment confidence.

H_{2b}: On a between-judge level, more frequently made judgments in the extreme ranges are associated with high judgment confidence regarding mathematics and verbal self-concepts.

H_3: The majority of teachers are overconfident in their judgments of their students' mathematics and verbal self-concepts.

H_{4a}: The target-specific confidence of teachers in their judgments of the mathematics and verbal self-concepts of their students is positively associated with the accuracy of these judgments.

H_{4b}: The judge-specific confidence of teachers in their judgments of the mathematics and verbal self-concepts of their students is positively associated with the accuracy of these judgments.

Method

Database and Sample

The present investigation is part of a larger research project concerning school and classroom climate, which addresses opportunities to improve student well-being. Participation in the research project was voluntary. In total, 21 public secondary schools (*Realschulen*[3]) in the German state of Bavaria were asked to participate; 19 of them took part. In these 19 schools, Grades 5–9 were included in the

sample. Because the majority of German secondary schools are organized according to the principle of having teachers specialize in specific school subjects, and because the teachers for main subjects teach several lessons per week in one class and therefore probably know their students best, data were collected from teachers of the two main subjects of mathematics and German.

Student data were collected in the classroom. Teacher questionnaires were distributed at the same time as student data were collected and then returned, individually, by the teachers themselves.

The present analysis was conducted on data of those teachers for whom full datasets were available. This comprises their assessments of the students' mathematics or verbal self-concepts, answers to queries on their judgment confidence and self-reports completed by the students on their academic self-concepts. Altogether, data were analyzed for a total of 96 teachers (47.9% women). On average, the teachers had been professionally active for 9.6 years, whereby this varied from 1 to 35 years. Furthermore, data from 1,388 students (45.5% girls) were included in the analyses. The average age of the students was 13.9 years ($SD = 1.79$). Each teacher conducted between seven and 32 student assessments ($M = 22.3$). Due to the previously mentioned principle of subject-specific teaching, mathematics teachers and German teachers are mutually exclusive, while a large proportion of the students were assessed by their mathematics as well as their German teachers. The 42 German language teachers (59.5% women) were associated with data for 1,036 students; the 54 mathematics teachers (38.9% women) were linked to data for 1,188 students.

Measurements

Verbal and mathematics self-concepts of students. The domain-specific self-concepts of the students were assessed using two scales taken from the Program for International Student Assessment (PISA) 2000 survey (see Kunter et al., 2003), which are specifically formulated for the subjects of German and mathematics (sample items: "I am a hopeless case in the subject of German"; "I am always good at math"). Each of the two scales comprised three items that were presented with a 4-point Likert-type answer scale ranging from 1 (*not true*) to 4 (*true*). In the PISA 2000 survey (see also Artelt, Baumert, Julius-McElvany, & Peschar, 2003), both scales show sufficient internal consistency (Cronbach's $\alpha =$ .90 and .83, respectively) and validity (criterion validity: for example, students with good performances had better self-concepts than students with weak performances). In the present study, the internal consistencies for both scales were satisfactory (German: Cronbach's $\alpha =$ .83; mathematics: $\alpha =$ .89).

Teacher judgments of students' self-concepts. Teachers were asked to judge the domain-specific self-concept of each of their students using the following question: "How does the

student evaluate his/her abilities in mathematics/German?" Judgments were assessed with a 4-point Likert-type answer scale ranging from 1 (*low*) to 4 (*high*).[4]

Judgment confidence of teachers. The confidence of teachers in their judgments of the students' self-concepts was assessed separately for each student. To this end, directly after the self-concept judgments, the teachers were asked to answer the question "How confident are you in this judgment?" A 4-point Likert-type answer scale was used ranging from 1 (*insecure*) to 4 (*secure*).

Results

To avoid overcomplexity, analyses were conducted separately for teacher judgments of mathematics self-concepts and teacher judgments of verbal self-concepts. This was also justified by the independence of mathematics and German self-concepts—in fact, they were slightly negatively correlated ($r = -.19$; $p < .001$).

Descriptive Analyses

Table 2 presents an overview of descriptive statistics for all variables. Judgment accuracy was operationalized with Spearman's correlation, calculated on a classroom-specific basis, between the domain-specific self-concepts indicated by the students and the corresponding judgments made by their teachers. These coefficients were subjected to a Fisher z transformation prior to all subsequent analyses to convert the nonequidistant coefficients into values which approximate the interval scale level. In addition to the means and standard deviations of the variables, percentile ranks were also reported to provide insight into the ranges and distributions of the variables. All variables showed a large range with the exception of mean teacher judgment confidence, which exhibited a medium range.

In a first step, tests were conducted to determine whether differences in judgment accuracy exist between the school subjects of mathematics and German. It was found that, on average, mathematics and German teachers assessed the self-concepts of their students comparably well ($z = 0.25$, $p = .80$, effect size $q = .05$; see Steiger, 1980).

Intraindividual Variations in Teachers' Confidence in Their Judgments

To determine the extent to which teachers are intraindividually different regarding their judgment confidence (H_1), the proportions of variances on a within-judge level and a between-judge level were calculated. The large proportions of intraindividual variances (i.e., on a within-judge level) found for both subjects indicate the substantial target specificity of judgment confidence (mathematics: 74.2%; German: 45.8%). It is worth noting here that the proportion of within-judge variance was, on a descriptive level, higher

TABLE 2. Psychometric Properties for All Variables (Subjects of Mathematics and German)

Variable	n	M	SD	Percentile				
				0	25	50	75	100
Mathematics								
Student self-concept (target)	1,188	2.61	0.94	1.00	2.00	2.67	3.33	4.00
Teacher target-specific judgment	1,188	2.57	0.76	1.00	2.00	3.00	3.00	4.00
Teacher target-specific judgment confidence	1,188	2.85	0.67	1.00	2.00	3.00	3.00	4.00
Teacher mean judgment confidence	54	2.87	0.39	1.50	2.63	2.85	3.09	3.70
Teacher judgment accuracy	54	0.27	0.30	−0.38	0.06	0.30	0.51	0.78
German								
Student self-concept (target)	1,036	2.73	0.72	1.00	2.33	2.83	3.33	4.00
Teacher target-specific judgment	1,036	2.49	0.76	1.00	2.00	3.00	3.00	4.00
Teacher target-specific judgment confidence	1,036	2.91	0.77	1.00	2.00	3.00	3.00	4.00
Teacher mean judgment confidence	42	2.91	0.53	1.00	2.61	2.83	3.24	3.96
Teacher judgment accuracy	42	0.22	0.27	−0.45	0.11	0.22	0.44	0.71

Note. Scale ranges for all measurements except judgment accuracy: 1–4. Judgment accuracy is quantified using classroom-wise calculated Spearman correlations between teacher judgments of the self-concept of the individual students and self-reported student self-concepts. Accounting for the unreliability of the self-concept measurements, mean judgment accuracies calculated using attenuation corrected coefficients were $\rho = .30$ for the subject of mathematics and $\rho = .27$ for the subject of German.

for the mathematics teachers than for the German teachers. Even when taking into account that error variance is usually located on a within-judge level, the large amounts of intraindividual fluctuations (particularly among the mathematics teachers) indicate that the confidence that teachers have in the judgments they make of individual students varies substantially in accordance with the student being assessed.

Relationship Between Judgment Confidence and Teacher Assessments

A regression analysis was conducted to determine whether teachers on a within-judge level are more secure in judgments of high or low self-concepts than when they assess students to have an average self-concept (H_{2a}). In the regression model, the confidence levels the teachers expressed in their student-specific judgments were regressed onto their student-specific judgments by applying linear and quadratic terms. The hierarchical structure of the data and the nonindependence of observations (nesting students within teachers) were accounted for by standard error correction. Model estimations were based on restricted maximum likelihood ratio parameters. Maximum likelihood ratio is deemed to be rather robust in estimating standard errors when faced with the lack of normal distribution; the present analyses were computed with Mplus 5.1 (Muthén & Muthén, 2007). The standardized linear coefficients proved to be significant: They amounted to $\beta_1 = .18$ ($p < .001$, $SE = .03$) for the subject of German and $\beta_1 = .14$ ($p < .001$, $SE = .04$) for the subject of mathematics. The quadratic term was also significant for both subjects: In the subject of German with $\beta_2 = .19$ ($p < .001$, $SE = .04$) and in the subject of mathematics with $\beta_2 = .28$ ($p < .001$, $SE = .04$). According to these

significant quadratic interrelations, mathematics teachers as well as German teachers are somewhat more confident in judging perceived high and low self-concepts than in judging average ones. The proportion of variance explained by the linear and quadratic terms was 10.6% for the mathematics teachers and 5.7% for the German teachers.

To test the hypothesis that teachers who vary more in their judgments are more confident in their judgments (between-judge level; H_{2b}), confidence judgments, aggregated within teachers, were correlated with the standard deviations of the self-concept judgments of each individual teacher. For German teachers this resulted in an insignificant correlation (r) of .07 ($p = .65$), for the mathematics teachers a significant correlation of .33 ($p = .02$) resulted from the analyses. However, the difference between both correlations was not significant ($z = 1.28$, $p = .20$, $q = .27$).

Overconfidence Bias

To determine whether a majority of the teachers are more confident in their judgments than is reflected in their accuracy (H_3), judgment accuracy (i.e., the teacher-specific correlation between teacher judgments and the self-reported self-concepts of the students) and mean judgment confidence had to be put into a relationship with one another. To accomplish this, normative cutoff values were necessary. In terms of judgment confidence, the scale midpoint value of the respective answer scale with the poles "insecure" and "secure" appeared to be a good approximation in establishing a cutoff point (scale midpoint value: 2.5). For judgment accuracy, a determination of this type was more difficult. If applying the results obtained from previous studies (e.g., Praetorius, Karst, Dickhäuser, & Lipowsky, 2011; Spinath,

TABLE 3. Frequencies of Combinations Between Accurate Versus Inaccurate Judgments and Low Versus High Judgment Confidences (Subjects of Mathematics and German)

Teacher mean judgment confidence	Teacher judgment accuracy		
	Low	High	Sum
Mathematics (n = 54)			
Low	13%(11%)	2%(4%)	15%(15%)
High	61%(50%)	24%(35%)	85%(85%)
Sum	74%(61%)	26%(39%)	100%(100%)
German (n = 42)			
Low	5%(5%)	2%(2%)	7%(7%)
High	83%(69%)	10%(24%)	93%(93%)
Sum	88%(74%)	12%(26%)	100%(100%)

Note. Judgment accuracies of $\rho > .50$ were classified as accurate judgments. Additionally calculated results for the less conservative criterion of $\rho > .40$ are presented in parentheses. Judgment confidence above the scale midpoint (2.5) was classified as high.

2005), then it is possible to classify a judgment accuracy of $\rho > .50$ as being high. To reduce potential errors resulting from the arbitrariness of setting such a normative cutoff, which could lead to mistakenly accepting the hypothesis, we decided to apply additionally a more moderate cutoff, whereby a teacher-specific correlation of $\rho > .40$ was classified as high judgment accuracy.

When applying the stronger judgment accuracy criterion of $\rho > .50$, we found that 61% of the mathematics teachers and 83% of the German teachers demonstrated an overconfidence bias (Table 3). Under the more moderate criterion, 50% of the mathematics teachers and 69% of the German teachers were still classified as members of this group. Thus, depending on the criterion applied and the academic subject considered, at least one half (and up to more than three quarters) of the teachers in the present sample were subject to the effects of an overconfidence bias.

Relationships Between Judgment Confidence and Judgment Accuracy

The hypothesis that the accuracy of individual teacher judgments and their confidence in these judgments are positively interrelated on a within-judge level (H_{4a}) was examined using a hierarchical multiple regression model with two levels and teachers' judgments of students' self-concepts as a dependent variable. In the first step (Model 1), the teacher judgments were regressed onto the domain-specific self-concepts reported by the students themselves. This model was used to determine the accuracy of the teacher assessments: The higher the regression coefficient, the higher the proportion of variance in the teacher judgments can be ascribed to the actual self-concepts of students. In the

second step, the target-specific judgment confidence of the teachers was included as a predictor variable (Model 2). Finally, in the third step, an interaction term was included as predictor (Model 3). The interaction term (students' self-concepts · teachers' judgment confidence) was formed by multiplying the students' domain-specific self-concepts (predictor) with the target-specific judgment confidence of the teachers (moderator variable). To avoid problems stemming from multicollinearity, the two independent variables were z transformed prior to multiplication. As standardized regression coefficients of interaction terms cannot be interpreted unambiguously (see Aiken & West, 1991), the unstandardized coefficients should be used for interpretation. The final model is represented by the following equation: *teacher judgments* = B_0 + B_1 · *students' self-concepts* + B_2 · *teachers' judgment confidence* + B_3 · *students' self-concepts · teachers' judgment confidence* + *e*.

Teacher confidence in their judgments could independently explain a proportion of variance in the judgments beyond that accounted for by students' domain-specific self-concepts—this was true for both mathematics and German teachers (Table 4). The interaction term (Model 3) turned out to be significant for mathematics teachers, but not for German teachers. By introducing the interaction term into the regression model, the proportion of explained criterion variance among the mathematics teachers could be significantly improved, albeit on a relatively small scale. Figure 1 illustrates the interaction effect for mathematics teachers. As depicted by the slopes of the regression lines, the relationship between student self-concepts and teacher judgments of these self-concepts was closer—and thus the accuracy was better—when the teacher was more confident in the judgment being made. In the case of mathematics teachers, who were highly insecure about their judgments, no relationship was found between teacher judgments and the actual self-concepts of their students.

The hypothesis that the overall accuracy of teacher judgments is positively interrelated with their judgment confidence on a between-judge level (H_{4b}) was examined on the basis of a correlation between the averaged judgment confidence for each teacher with their averaged judgment accuracy. For the subject of German this resulted in a nonsignificant correlation of $r = -.14$ ($p = .39$), however, the correlation $r = .30$ ($p = .03$) for the subject of mathematics was significant.

Discussion

The aims of the present study were first to determine how confident teachers are in their judgments of their students' domain-specific self-concepts in the subjects of mathematics and German and, second, whether these judgment confidences are associated with judgment accuracy. The hypotheses were derived from research in the field of social psychology because, to the best of our knowledge, no prior

TABLE 4. Hierarchical Multiple Regression Analyses Predicting Teacher Judgments of Student Self-Concepts and Teacher Target-Specific Judgment Confidences (Subjects of Mathematics and German)

	Model 1					Model 2					Model 3				
Variable	B	SE	95% CI	β	ΔR^2/Total R^2	B	SE	95% CI	β	ΔR^2/Total R^2	B	SE	95% CI	β	ΔR^2/Total R^2
Mathematics															
Student SC (target)	.25	.04	[0.17, 0.33]	.31*	10***/.10	0.24	.04	[0.16, 0.32]	.30*		.24	.04	[0.16, 0.32]	.30*	
Teacher target-specific JC						0.19	.04	[0.11, 0.27]	.17*	.02***/.12	.20	.04	[0.12, 0.28]	.18*	
SC·JC											.12	.03	[0.06, 0.18]	.17*	.03***/.15
German															
Student SC (target)	.21	.05	[0.11, 0.31]	.20*	.04***/.04	0.20	.05	[0.10, 0.30]	.19*		.20	.04	[0.12, 0.28]	.19*	
Teacher target-specific JC						0.14	.04	[0.06, 0.22]	.14*	.02***/.06	.14	.04	[0.06, 0.22]	.14*	
SC·JC											.02	.03	[−0.04, 0.08]	.02	.00/.06

Note. The analyses are based on data collected from 1,188 students and 54 teachers for the subject of mathematics as well as 1,036 students and 42 teachers for the subject of German. CI = confidence interval; SC = self-concept; JC = judgment confidence.
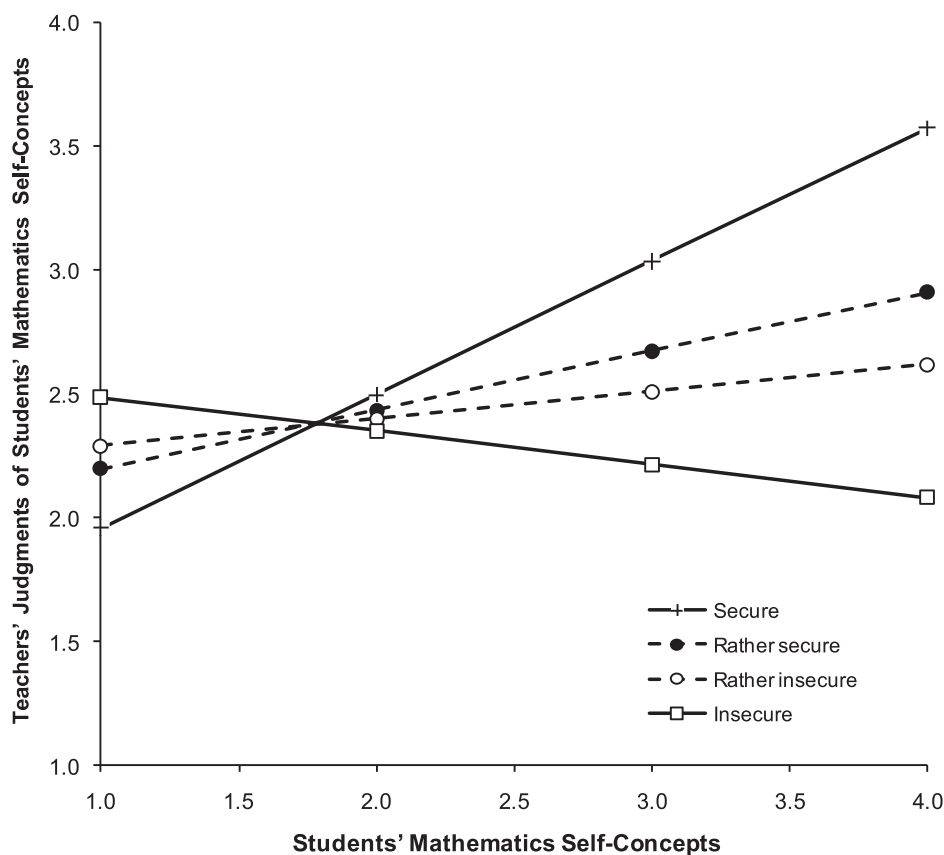*p < .05. ***p < .001.

71

FIGURE 1. Relationship between student self-concepts and teacher judgments of student self-concepts in the subject of mathematics depending on teacher target-specific confidence in their judgments.

research findings have been published on the topic of teacher confidence in their judgments of student characteristics.

What Makes a Teacher Feel Confident in His or Her Judgment?

We first tested the presumption that teachers demonstrate intraindividual variations regarding their judgment confidence. As expected, these intraindividual variations were confirmed in the analyses. This variability justifies the consideration of teachers' confidence in their judgments on the within-judge level as a relevant unit of analysis made in the subsequent steps of the analyses.

Subsequently, focus was placed on the relationship between extreme (both high and low) judgments of their students' characteristics and judgment confidence by conducting regression analyses with linear and quadratic terms. The hypothesis that teachers express higher levels of confidence concerning judgments of high and low self-concepts than they do about judgments which are located in the midrange could be confirmed. The significant linear term indicated a general response tendency: Individuals who tend to agree on

one item tend to agree on another item as well. This would imply that content-independent reasons caused the relationship between teacher judgments and their judgment confidence. However, this effect cannot be the only explanation for the result as the quadratic term was also revealed to be significant. This quadratic term indicates that the cues which teachers use as indicators for high and low domain-specific self-concepts are perceived as more salient and, therefore, more conclusive than the cues teachers understand as indicators of average self-concept levels. The actual cues drawn on by teachers in such judgment situations would be a worthwhile subject for follow-up investigations. Reflecting on the practical implications of these results, it is possible to conclude that higher confidence in judgments of extreme values, in comparison with midrange judgments, is appropriate and functional, as students with very low self-concepts are in particular need of interventions (see Marsh et al., 2003).

In the next step, the investigation of relationships between extreme judgments of self-concepts and judgment confidence was expanded from a within-judge level to a

between-judge level. Here, too, the hypothesis that extreme judgments are associated with higher levels of confidence was confirmed for the subject of mathematics, but not for the subject of German. For the subject of mathematics this confirms the assumption that persons have a tendency to make assessments in the middle of the available range when they feel unsure about the target presented to them. In his groundbreaking article on the measurement of judgment accuracy, Cronbach (1955) concluded, "it now appears that the teacher who is poorly informed regarding the unique patterns of his [or her] students should probably treat them by a standard pattern of instruction, carefully fitted to the typical student" (p. 183). Because German teachers are not, and mathematics teachers are only partially, aware of whether they are capable of making good or poor diagnoses, a point should be made to enhance judgment accuracy as well as adequate judgment confidence. As expressed by Cronbach, efforts to implement adaption in classrooms can otherwise have negative effects: "Differentiation is harmful if the extent of adaption or differentiation exceeds the amount justified by the accuracy of social perception" (p. 183). However, as the present study is not an experimental one, alternative explanations of the aforementioned effect cannot be ruled out completely. Therefore, future experimental research would be desirable.

The More Confident, the More Accurate? Relationships Between Judgment Confidence and Judgment Accuracy

It has been suggested that most teachers are more confident in the judgments they make than would be warranted when the accuracy of these judgments is examined. The analyses conducted here confirm this suspicion. This effect can be accounted for by the fact that persons exhibit a substantial over-reliance on the strength of the cues available to them (Griffin & Tversky, 1992), also referred to as the availability heuristic. In the lens model developed by Brunswik (1952), cues are broken down into three groups: valid, invalid, and irrelevant cues. To enable a more detailed understanding of the distortions that are present in teachers' judgments of their students' self-concepts, future researchers should first determine what types of cues are valid for the judgment of self-concepts. Such an overview is presently not available. Second, there is a need for conducting investigations that determine which cues teachers actually rely on.

It was furthermore presumed that judge-specific confidence and target-specific confidence is positively correlated with judgment accuracy. The results of our analyses indicate that teachers are, at best, only partially aware of whether their diagnoses are correct: For the school-subject of German, a correlation between judgment accuracy and judgment confidence could be found on neither a between-judge level nor a within-judge level. In other words, teachers in the subject of German do not seem to be aware of whether they are capable of adequately diagnosing student characteristics. This situation is somewhat different for the subject

of mathematics. Here, a moderate correlation was detected between judgment accuracy and judgment confidence on a between-judge level. This was confirmed on a within-judge level: In the subject of mathematics, teachers differentiate among their students with regard to their judgment confidence. In contrast to the findings for the subject of German, these differentiations were correlated with actual judgment accuracy: The incremental validity of the interaction term found between mathematics self-concept and judgment confidence with respect to the outcome of teacher judgments of student self-concepts corresponds, according to Cohen (1992), to a small effect. Nevertheless, magnitudes of this level are typical for interaction effects (see Chaplin, 1991). In considering previous investigations, for example studies conducted in the field of social psychology, inconsistent findings also surface: On one hand, some studies have reported a correlation between judgment confidence and judgment accuracy on a between-judge level (Smith et al., 1991; Vallone et al., 1990)—the same pattern of results we found for teacher judgments in the subject of mathematics. A correlation of this type is desirable in that it signifies that teachers tend to be—at least partly—aware of being able (or unable) to rely on the judgments they make. This plays a significant role in the decision to either adapt subsequent actions to this judgment or to consider relying on alternative instructional options. On the other hand, there are findings that indicate that there is no correlation between judgment confidence and judgment accuracy (such as Ames et al., 2010; Ickes, 1993; Patterson et al., 2001). This lack of a relationship is evident among teachers for the subject of German. Funder (1999) explained this pattern in the following terms:

> A good judge is likely to know that he or she is good, but there is nothing to stop a poor judge from also believing himself or herself to be a good judge. [. . .] Thus, it is not surprising that self-assessments of judgmental ability are seldom found to be correlated with actual judgmental success. (p. 142)

This leads to the following conclusion: If judgment accuracy is to encouraged among a group of teachers (here, specifically, for the subject of German), the teachers must first be made aware of how well they can judge the characteristics of their students. Therefore, training in judgment accuracy will only be fruitful when teachers themselves recognize a need for support. Encouragement of this type is necessary: Motivating students to learn is an important task, one that must be executed by teachers nearly every day. Our results indicate that, to some extent, the wrong students are being supported and that German teachers, and in some cases also mathematics teachers, are confident that their decisions in designating specific motivational practices to specific students are appropriate.

However, how can teachers discern whether they are good diagnosticians? We have not learned much about the indicators that tell us how accurate diagnoses are to date. According to Funder (1999), good diagnosticians need implicit or explicit information on the characteristics they are assessing

and how these are manifested in behavior; moreover, practical experience and feedback are particularly effective in forming knowledge of assessing student characteristics. Unfortunately, very few teachers are given feedback on their judgment accuracy—this fact provides a starting point for the promotion of judgment accuracy.

Unique Patterns? The Generalization of the Findings Across the Subjects of Mathematics and German

To enable a generalization of the findings, two academic subjects (mathematics and German) were included in the present investigation. Our findings for these two subjects converged with respect to the substantial target-specificity of judgment confidence, the stronger confidence in judgments of high and low self-concepts than in judgments of average self-concepts on a within-judge level as well as the predominance of an overconfidence bias. We therefore observe a sufficient generalizability of these findings. However, the results obtained also indicate that mathematics teachers are—at least partially—aware of their judgment accuracy, but German teachers are not. These domain differences point to a limited generalizability of this specific finding. One possible post hoc explanation for the difference between the two subjects may be that in mathematics, due to the relatively explicit feedback an individual receives, students may be able to form more veridical self-concepts of their mathematics capabilities. On the other hand, in the subject of German, there is a much wider scope for the interpretation of an individual's performances, therefore verbal self-concepts may deviate more strongly from actual capabilities than for the subject of mathematics. Should this be the case, judging an individual's verbal self-concept would be a significantly more difficult task than judging his or her mathematics self-concept. German teachers would then be more likely to make false judgments, and simultaneously would not be aware of this increased difficulty (possibly due to a lack of comparative standards). Whether this explanation is appropriate can only be determined by future empirical research.

Limitations and Directions for Further Research

It is important to note some limitations of the present research that may influence the interpretation of the findings as well as provide reasonable directions for future research. First, our study was restricted to mathematics and German teachers working in the German public school system. Thus, it is possible that the results cannot be generalized to other school subjects, or to school systems unlike the German model. Second, statements can only be made on the accuracy of and confidence in judgments concerning the domain-specific self-concepts held by students. Whether, for example, the pattern of results is the same for judgments of student achievement cannot be clarified. Third, as the study was cross-sectional, causality could not be assessed. Therefore it remains unclear whether confidence in judgments

exerts an influence on the judgments themselves and their accuracy, or whether the degree of judgment accuracy influences the confidence in an individual's judgments. Four, both teacher measures used in the present study contained only one single item. Although such single-item measures are commonly used in judgment accuracy research, they can be challenged as they are suspected to have low reliabilities, thus lowering the relationships between variables. However, some studies have pointed out that this is not necessarily the case (Bergkvist & Rossiter, 2007; Loo, 2002; Wanous, Reichers, & Hudy, 1997), for example when the construct under consideration is rather narrow and can be interpreted unambiguously. Whether single-item measures are sufficient for investigating teacher judgments and judgment confidence should be investigated in further studies.

Based on the confidence expressed by teachers in their judgments, judgment accuracy can be partially explained, at least for the subject of mathematics. This is an initial finding that may be fundamental for future research. Future researchers should place a strong emphasis on the theoretically postulated effects of judgment confidence described above: Is it truly the case that teachers who are confident in the judgment they make of learners' prerequisites are more likely to implement them in the classroom? Also, the question of the determinants of judgment confidence is still pending. Why is it that some teachers are confident in the judgments they make and others are not? And why is a teacher who is confident in the judgment made on one student insecure when judging another? Here further research is clearly needed.

NOTES

1. To explain this unexpected result, Marsh (1986) developed the internal/external frame of reference model (I/E model). According to this model, the formation of self-concepts is dependent on two information sources: social comparisons (e.g., comparisons of an individual's own mathematic performance with those of his or her classmates) and dimensional comparisons (e.g., comparison of an individual's own mathematic performance with his or her own verbal performance). The influence of both comparisons on the self-concept is in the opposite direction of one another (positive for social and negative for dimensional comparisons). Thus, their influences offset one another that results in nonsignificant correlations between the mathematic and the verbal self-concept.

2. The subject of German (native language for most of the students) covers mainly literature and grammar.

3. In the German public school system, following Grade 4, students are allocated to one of the following three school types on the basis of academic achievement: The curriculum in a *Hauptschule* lasts for a period of 5 years, the curriculum in a *Realschule* lasts 6 years and in a *Gymnasium*—depending on federal state—students study for 8 or 9 years. The first two school types train students for future occupations (apprenticeship), and the third serves as preparation for university studies.

4. In judgment accuracy research, teacher judgments concerning students' characteristics are usually surveyed with single-item scales (for an overview, see Südkamp, Kaiser, & Möller, 2012). To ensure comparability to this research and for economic reasons we decided to gather only one judgment per student, too. Thus, reliability cannot be estimated. Nevertheless, as judgment accuracy and instructional effectiveness are interrelated (Helmke & Schrader, 1987; cf. Anders et al., 2010), predictive validity of the teacher judgment measure can be assumed. Analogously to the teacher judgments we also collected only one item for the confidence judgments. Reporting reliability estimates for the confidence judgments is thus not possible, too.

REFERENCES

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Ames, D. R., Kammrath, L. K., Suppes, A., & Bolger, N. (2010). Not so fast: The (not-quite-complete) dissociation between accuracy and confidence in thin-slice impressions. *Personality and Social Psychology Bulletin, 36,* 264–277. doi:10.1177/0146167209354519

Anders, Y., Kunter, M., Brunner, M., Krauss, S., & Baumert, J. (2010). Diagnostische Fähigkeiten von Mathematiklehrkräften und ihre Auswirkungen auf die Leistungen ihrer Schülerinnen und Schüler [Mathematics teachers' diagnostic skills and their impact on students' achievements]. *Psychologie in Erziehung und Unterricht, 57,* 175–193. doi:10.2378/peu2010.art13d

Artelt, C., Baumert, J., Julius-McElvany, N., & Peschar, J. (2003). *Learners for life. Student approaches to learning. Results from PISA 2000.* Paris, France: OECD.

Baker, S. F. (2010). *Calibration analysis within the cognitive and personality domain: Individual differences in confidence, accuracy, and bias* (Doctoral dissertation). University of Southern Queensland, Australia. Retrieved from http://eprints.usq.edu.au/8856/2/Baker_2010_whole.pdf

Begeny, J. C., Eckert, T. L., Montarello, S. A., & Storie, M. S. (2008). Teachers' perceptions of students' reading abilities: An examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly, 23,* 43–55. doi:10.1037/1045-3830.23.1.43

Bergkvist, L., & Rossiter, J. R. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research, 44,* 175–184. doi:10.1509/jmkr.44.2.175

Boehnke, K., Silbereisen, R. K., Reynolds, C. R., & Richmond, B. O. (1986). What I think and feel—German experience with the revised form of the children's manifest anxiety scale. *Personality and Individual Differences, 7,* 553–560. doi:10.1016/0191-8869(86)90132-7

Brunswik, E. (1952). The conceptual framework of psychology. In *International encyclopedia of unified science* (Vol. 1). Chicago, IL: University of Chicago Press.

Chaplin, W. F. (1991). The next generation in moderation research in personality psychology. *Journal of Personality, 59,* 143–178. doi:10.1111/j.1467-6494.1991.tb00772.x

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159. doi:10.1037/0033-2909.112.1.155

Corno, L., & Snow, R. E. (1986). Adapting teaching to individual differences among learners. In M. C. Wittrock (Ed.), *Third handbook of research on teaching* (pp. 605–629).Washington, DC: American Educational Research Association.

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin, 52,* 177–193.

DePaulo, B. M., Charlton, K., Cooper, H., Lindsay, J. M., & Muhlenbruck, L. (1997). The accuracy-confidence correlation in the detection of deception. *Personality and Social Psychology Review, 1,* 346–357. doi:10.1207/s15327957pspr0104_5

Edelenbos, P., & Kubanek-German, A. (2004). Teacher assessment: the concept of "diagnostic competence." *Language Testing, 21,* 259–283. doi:10.1191/0265532204lt284oa

Fox, S., Bizman, A., Hoffman, M., & Oren, L. (1995). The impact of variability in candidate profiles on rater confidence and judgments regarding stability and job suitability. *Journal of Occupational and Organizational Psychology, 68,* 13–23.

Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception.* San Diego, CA: Academic Press.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24,* 411–435.

Helmke, A., & Schrader, F.-W. (1987). Interactional effects of instructional quality and teacher judgment accuracy on achievement. *Teaching and Teacher Education, 3,* 91–98. doi:10.1016/0742-051X(87)90010-2

Hoge, R. D., & Coladarci, T. (1989). Teacher based judgments of academic achievement: A review of the literature. *Review of Educational Research, 59,* 297–313. doi:10.2307/1170184

Ickes, W. (1993). Empathic accuracy. *Journal of Personality, 61,* 587–610. doi:10.1111/j.1467-6494.1993.tb00783.x

Klayman, J., Soll, J. B., Juslin, P., & Winman, A. (2006). Subjective confidence and the sampling of knowledge. In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition* (pp. 153–182). Cambridge, UK: Cambridge University Press.

Kunter, M., Schümer, G., Artelt, C., Baumert, J., Klieme, E., Neubrand, M., & Weiß, M. (2003). *PISA 2000: Dokumentation der Erhebungsinstrumente [PISA 2000: Documentation of the survey instruments]* (Vol. 72). Berlin, Germany: Max-Planck-Institut für Bildungsforschung.

Lichtenstein, S., Fischhoff, B., & Philips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgments under uncertainty: Heuristics and biases* (pp. 306–334). New York, NY: Cambridge University Press.

Loo, R. (2002). A caveat on using single-item versus multiple-item scales. *Journal of Managerial Psychology, 17,* 68–75. doi:10.1108/026839-40210415933

Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal, 23,* 129–149. doi:10.3102/00028312023001129

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science, 1,* 133–162. doi:10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., Craven, R. G., & McInerney, D. M. (2003). International advances in self research: Speaking to the future. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), *International advances in self research* (Vol. 1, pp. 3–14). Greenwich, CT: Information Age.

Metcalfe, J. (1998). Cognitive optimism: Self-deception or memory-based processing heuristics? *Personality and Social Psychology Review, 2,* 100–110. doi:10.1207/s15327957pspr0202_3

Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external Frame of reference model of academic achievement and academic self-concept. *Review of Educational Research, 79,* 1129–1167. doi:10.3102/0034654309337522

Muthén, L. K., & Muthén, B. O. (2007). *Mplus user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.

Myford, C. M., & Wolfe, E. W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement, 4,* 386–422.

Patterson, M. L., Foster, J. L., & Bellmer, C. G. (2001). Another look at accuracy and confidence in social judgments. *Journal of Nonverbal Behavior, 25,* 207–219.

Praetorius, A.-K., Karst, K., Dickhäuser, O., & Lipowsky, F. (2011). Wie gut schätzen Lehrer die Fähigkeitsselbstkonzepte ihrer Schüler ein? Zur diagnostischen Kompetenz von Lehrkräften [How teachers rate their students: On teachers' diagnostic competence regarding the academic self-concept]. *Psychologie in Erziehung und Unterricht, 58,* 81–91. doi:10.2378/peu2011.art30d

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77,* 454–499. doi:10.3102/0034654307310317

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research, 46,* 407–441. doi:10.3102/00346543046003407

Smith, H. J., Archer, D., & Costanzo, M. (1991). "Just a hunch": Accuracy and awareness in person perception. *Journal of Nonverbal Behavior, 15,* 3–18.

Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz [Accuracy of teacher judgments on student characteristics and the construct of diagnostic competence]. *Zeitschrift für Pädagogische Psychologie, 19,* 85–95. doi:10.1024/1010-0652.19.12.85

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87,* 245–251.

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology, 104,* 743–762.

Vallone, R. P., Griffin, D. W., Lind, S. L., & Ross, L. (1990). Overconfident prediction of futures actions and outcomes by self and others. *Journal of Personality and Social Psychology, 58,* 582–592.

Wanous, J. P., Reichers, A. E., & Hudy, M. J. (1997). Overall job satisfaction: How good are single-item measures? *Journal of Applied Psychology, 82,* 247–252.

West, R. F., & Stanovich, K. E. (1996). The domain specificity and generality of overconfidence: Individual differences in performance estimation bias. *Psychonomic Bulletin and Review, 4,* 387–392.

AUTHORS NOTE

Anna-Katharina Praetorius, M.A., is affiliated with the Department of Psychology at the University of Augsburg in Germany. Her research interests include instructional quality, teacher competencies and self-concept of students.

Dipl.-Päd. Valérie-Danielle Berner is affiliated with the Department of Psychology at the University of Augsburg in Germany. Her research interests include learning and motivation of students, school and family integration, heterogeneity and school achievement.

Prof. Dr. Horst Zeinz is affiliated with the Institute of Education at the University of Muenster in Germany. His research interests include school development, implementation of innovations in schools and teacher cognition.

Prof. Dr. Annette Scheunpflug is affiliated with the Department of Education at Friedrich-Alexander-University Erlangen-Nuremberg in Germany. Her research interests include global aspects of education, comparative education and quality of education.

Prof. Dr. Markus Dresel is affiliated with the Department of Psychology at the University of Augsburg in Germany. His research interests include learning motivation and self-regulated learning in various educational contexts, gender issues in education, instructional research and learning with digital media.