A central aspect of teachers' professional knowledge and competence is the ability to assess students' achievements adequately. Giving grades and marks is one prototypical task in this context. Besides giving grades, assessments for school placements or tracking decisions belong to these tasks. Relevant students' characteristics which influence teachers' assessments do not only involve academic achievement but also students' responses to different task demands as well as non-academic characteristics such as learning motivation or school anxiety.

Closely associated with the investigation of teachers' assessment competences and, more specifically, the investigation of conditions associated with high quality of assessments is the development and evaluation of teacher training programs to improve professional competences. In recent years, there has been considerable progress in the domain of professional teacher training; however, only a very limited number of studies are dedicated to the question to what extent training programs might offer valuable approaches to improve the quality of assessments and to implement high assessment competences.

Another important field which is closely related to teachers' competences concerns the question how teachers' professional development is linked to students' learning and learning outcomes. In recent years, the societal demand for evidence that teachers' professional development will result in improved student learning outcomes is increasing.

This volume brings together questions on assessment, training, and learning in the professional development of teachers which have not been fully discussed yet. The identification of these research gaps was the reason for dedicating a series of lectures given at the University of Luxembourg 2012 to the topic of professionalization of teachers in these domains. Therefore, this book contains contributions from outstanding international scholars in different academic disciplines to present ideas about open research questions concerning the domains of assessment, training, and learning in the professional development of teachers.
Teachers’ Professional Development
One characteristic of modern societies is that they are likely to assign their social problems to education. Arising in the specific context of the late eighteenth century, this ‘educational reflex’ paved the way for education to become an important social factor on regional, national and global scales. Witnesses for this upswing are for instance the expansion of compulsory schooling, the state organization and tertiarization of teacher education and thus the introduction of education departments in the universities.

However, in contrast to the social artefact of modern societies – pluralism in languages, cultures, values, and customs –, education research seems in many respects still committed to ideas of unity or uniformity: For instance, the global standardization movement fosters uniformity in curriculum and content to serve the purpose of dominant global evaluation schemes, which in turn are based on the idea of human cognition as an immutable arrangement of mental processes with regard to learning. Moreover, critics of these developments often argue with arguments and convictions that can be traced back to the time when the education sciences emerged in the context of the cultural and political idea of the uniform national state.

Obviously, today’s education research often operates using concepts that are derived from ideas of unity and uniformity in order to tackle the challenges of cultural and linguistic plurality in the context of democratic societies. This is both a paradox and an occasion to reflect upon the present and future role of education research in the context of modern societies in three attempts: *Education Systems in Historical, Cultural, and Sociological Perspectives* (Vol. 1); *Multilingualism and Multimodality: Current Challenges for Educational Studies* (Vol. 2); *Teachers’ Professional Development: Assessment, Training, and Learning* (Vol. 3).
Teachers’ Professional Development

Assessment, Training, and Learning

Edited by

Sabine Krolak-Schwerdt
University of Luxembourg, Luxembourg

Sabine Glock
University of Luxembourg, Luxembourg

and

Matthias Böhmer
University of Luxembourg, Luxembourg

SENSE PUBLISHERS
ROTTERDAM/BOSTON/TAIPEI
# TABLE OF CONTENTS

The Future of Education Research: Introduction to the Series of Three Volumes vii

*Daniel Tröhler*

Introduction 1

*Sabine Krolak-Schwerdt, Sabine Glock & Matthias Böhmer*

Teachers’ Judgments of Students’ Academic Achievement: Results From Field and Experimental Studies 5

*Anna Südkamp, Johanna Kaiser & Jens Möller*

Accuracy of Teacher Judgments: When and for What Reasons? 27

*Cordula Artelt & Tobias Rausch*

Improving Teachers’ Judgments: Obtaining Change Through Cognitive Processes 45

*Ineke Pit-ten Cate, Sabine Krolak-Schwerdt, Sabine Glock & Maria Markova*

Training Prospective Teachers in Educational Diagnostics 63

*Monika Trittel, Mara Gerich & Bernhard Schmitz*

Teacher Learning and Professional Development 79

*Jan D. Vermunt*

Professional Development Across the Teaching Career: Teachers’ Uptake of Formal and Informal Learning Opportunities* 97

*Dirk Richter, Mareike Kunter, Uta Klusmann, Oliver Lüdtke & Jürgen Baumert*

Epistemological Beliefs and Students’ Adaptive Perception of Task Complexity 123

*Rainer Bromme, Stephanie Pieschl & Elmar Stahl*

The Validity of Predictors of Academic and Vocational-Training Achievement: A Review of the Literature 153

*Florian Klapproth & Paule Schaltz*

Index 167
One characteristic of modern societies is that they are likely to assign their social problems to education. Arising in the specific context of the late eighteenth century, this ‘educational reflex’ paved the way for education to become an important social factor on local, regional, national and global scales. Witnesses for this upswing are, for instance, the expansion of compulsory schooling, the state organization and tertiariization of teacher education and thus the introduction of educational departments in the universities, the introduction of certificates for both students and teachers. However, in contrast to the social artefact of modern societies – pluralism in languages, cultures, values, and customs – the educational sciences seem in many respects still committed to ideas of unity or uniformity: For instance, the global standardization movement fosters uniformity in curriculum and content to serve dominant global evaluation schemes. These schemes in turn are based on the idea of human cognition as an immutable arrangement of mental processes with regard to learning. And the critics of these developments often argue with motives, arguments, and convictions that can be traced back to the time when the educational sciences emerged in the context of the cultural and political idea of the uniform (and of course superior) national state. In other words: Today, often the education sciences operate using concepts that are derived from ideas of unity and uniformity in order to tackle the challenges of cultural and linguistic plurality in the context of democratic societies. This obviously is both a paradox and an occasion to reflect about the present and future role of the educational sciences in the context of modern societies.

With over 40% of inhabitants not having Luxembourgish passports, Luxembourg is a multinational and thus a multilingual and multicultural society. With its three official languages Luxembourgish, German, French, and with Portuguese as first language of nearly 20% of the inhabitants, it is also a multilingual society. Against this background, Luxembourg is predestined to evaluate the ‘educational reflex’ mentioned above, the assigning of social problems to education. The University of Luxembourg responded to this desideratum by making ‘Education and Learning in Multilingual and Multicultural Contexts’ a Research Priority in the frame of the current four-year-plan (2010-2013).

One particular challenge of this research priority is the self-reflection or critical self-evaluation of the educational sciences in the context of the social expectations concerning education. Therefore, one of the major aims of “Education and Learning in Multilingual and Multicultural Contexts” was to assess the future of educational
research with outstanding international scholars. The 2010-2013 lecture series “The Future of Education Research” is an integral part of this research priority. Here the international discussion is not restricted to questions regarding technical feasibility and methods of educational ambitions. Self-reflection or critical self-evaluation meant precisely refraining from compliant adoptions of research desiderata defined by stakeholders of political, cultural, religious, or developmental institutions and being engaged in the (self-) critical assessment of the legitimacy and general feasibility of educational desiderata, that is of social expectations emerging from the educational reflex. Education research was defined not simply as a service towards fulfilling social expectations but like any other academic discipline a field in which its actors, the researchers, define the appropriateness of its research agenda – research questions and methods – in the realm of their peers.

With these premises, the future of education research is defined to be international, self-reflexive, and interdisciplinary and to include a broad range of traditional academic disciplines such as the education sciences in the narrower sense, psychology, sociology, linguistics, history, political sciences, cognitive sciences, and neurology sciences. And it is meant to focus on the macro, meso and micro levels of education questions and problems analytically, empirically, and historically. The invited international colleagues addressed their respective scholarship to the topic under consideration, the future of education research, in one of three lecture series at the University of Luxembourg from 2010 to 2013. In accordance with the interdisciplinary approach, the relevant questions were not clustered around traditional disciplines but around several focal points, resulting in this series of the following three volumes to be published between 2011 and 2014:

– *Education Systems in Historical, Cultural, and Sociological Perspectives* (Vol. 1)
– *Multimodality and Multilingualism: Current Challenges for Educational Studies* (Vol. 2)
– *Teachers’ Professional Development: Assessment, Training, and Learning* (Vol. 3)

We greatly appreciate the support of the University of Luxembourg and extend thanks for the opportunity to establish a Research Priority dedicated to “Education and Learning in Multilingual and Multicultural Contexts,” within which the lecture series “The Future of Education Research” is being held. We are grateful to all the excellent international scholars participating in this research discussion. And last but not least, we sincerely thank Peter de Liefde of Sense Publishers for his support of this series and for giving us, by means of publication, the opportunity to open up this discussion on a more global level.

Walferdange, Luxembourg, August 2011

Daniel Tröhler, Head of the Research Priority
“Education and Learning in Multilingual and Multicultural Contexts”,
University of Luxembourg
INTRODUCTION

The conditions and consequences of societal change are the focal point of current debates concerning professional development of actors in the educational domain, most notably teachers. The major goal is to make teacher education a profession with a research base and formal body of knowledge and to ensure that teachers are fully prepared in accordance with professional standards. In recent years, the teaching profession has begun to identify and develop the knowledge base that will frame the education curriculum.

A central aspect of teachers’ professional knowledge and competence is the ability to assess students’ achievements adequately. Giving grades and marks is one prototypical task in this context. Besides giving grades, assessments for school placements or tracking decisions belong to these tasks. Relevant students’ characteristics which influence teachers’ assessments do not only involve academic achievement but also students’ responses to different task demands as well as non-academic characteristics such as learning motivation or school anxiety. Teachers’ assessments have substantial relevance for individual students, and consequently, high competence in assessing students correctly is seen as a key skill for teachers.

Closely associated with the investigation of teachers’ assessment competences and, more specifically, the investigation of conditions associated with high quality of assessments is the development and evaluation of teacher training programs to improve professional competences. In recent years, there has been considerable progress in the domain of professional teacher training; however, only a very limited number of studies are dedicated to the question to what extend training programs might offer valuable approaches to improve the quality of assessments and to implement high assessment competences.

Another important field which is closely related to teachers’ competences concerns the question how teachers’ professional development is linked to students’ learning and learning outcomes. In recent years, the societal demand for evidence that teachers’ professional development will result in improved student learning outcomes is increasing. Current theorizing postulates a long chain of intermediate steps and variables which links teachers’ professional development to students’ learning. For instance, teachers’ beliefs about (good) teaching methods and students’ beliefs about learning might constitute such intermediate variables. There is, however, little research which covers the whole causal chain.
Taken together, questions on assessment, training, and learning in the professional development of teachers have not been fully discussed. The identification of these research gaps was the reason for dedicating the third round of lectures in the University of Luxembourg’s 2010-2013 lecture series “The Future of Educational Research” to the topic of professionalization of teachers in these domains. It was therefore our privilege to invite outstanding international scholars in different academic disciplines to present ideas about open research questions concerning the domains of assessment, training, and learning in the professional development of teachers.

In correspondence to these thematic foci, the first part of this volume is concerned with teachers’ assessment competences. Conceptualizing assessments as judgments about students’ characteristics, recent approaches frequently take a view on the quality of assessments as “judgment accuracy”. Correspondingly, the first part of this volume is concerned with teachers’ judgments of students’ academic achievement and judgment accuracy.

Südkamp, Kaiser, and Möller present a comprehensive model of teacher-based judgments of students’ academic achievement. In line with the model, the authors present the results of a meta-analysis of 75 field studies reporting correlational data on the relationship between teachers’ judgments of students’ academic achievement and students’ performance on a standardized achievement test. As to teacher characteristics, main results are that teachers who are informed about the content of a test prior to their judgment of students’ academic achievement perform better than uninformed teachers. Moreover they found that the congruence between the teachers’ judgment task and the achievement test administered to students is related to teacher judgment accuracy, with higher congruence being associated with higher accuracy levels. To further analyze the causal role of these variables, the authors recommend the combination of two research approaches, that is, an experimental approach and the validation of its results by field studies.

Artelt and Rausch review the empirical findings on teachers’ judgment accuracy and discuss potential moderators. Beyond students’ characteristics, task characteristics contribute to teachers’ accuracy, as teachers’ judgment accuracy is increasing with the correspondence between (1) the judgment scale and the test scale and (2) the judgment domain and the test domain. Artelt and Rausch present findings from the BiKS research group in Bamberg and provide evidence for teachers’ global achievement judgments of a particular domain being more accurate than their task-specific judgments. Task specific judgments relate to students’ ability to solve particular items in a test, while global judgments have no relations to particular tests. In the remainder of the chapter, the authors discuss necessary conditions for high judgment accuracy to occur. Among other variables, teachers’ stereotypical expectations contribute to judgment accuracy.

Pit-ten Cate, Krolak-Schwerdt, Glock, and Markova focus on the change of cognitive processes in order to overcome expectation biases and to improve teachers’ judgment accuracy. Frequently, the use of stereotypical knowledge about students is
INTRODUCTION

discussed as a potential source of judgment biases. However, the use of stereotypical knowledge in judgments is not inevitable, as there are techniques to overcome stereotypical biases such as the training of stereotype suppression. Furthermore, the motivation of the person making a judgment also plays a pivotal role in preventing teachers from reliance on stereotypical knowledge. It is demonstrated that one possible mechanism to increase accuracy motivations among teachers is to make them highly accountable for their judgment. A third aspect to increase teachers’ judgment accuracy consists of using of statistical prediction rules for judgment formation. These are formal decision rules on how information which has proven to be diagnostic for the judgment task should be weighted and integrated into a judgment.

After having discussed the conditions necessary for sufficient teachers’ judgment accuracy and how judgment accuracy could be improved, the second part of this volume deals with the development and evaluation of teacher training programs. Trittel, Gerich, and Schmitz are concerned with the evaluation of a program developed to train prospective teachers in assessment competence. They introduce a hands-on seminar which is based on a process model of teachers’ assessment competence. This program provides prospective teachers with theoretical knowledge about educational assessments and corresponding quality criteria as well as with knowledge about judgment biases. Additionally, prospective teachers are introduced to assessment instruments and methods. The application of these instruments and methods is practiced in lessons where prospective teachers are also prepared to plan supportive measures relying on the preceding assessment process. First results demonstrate the usefulness of this program in training prospective teachers’ assessment competence.

In the paper of Vermunt, teacher education and professional development is suggested to be causally related to student learning outcomes. Drawing on this chain model of teacher education and student learning outcomes Vermunt discusses the different components of student learning as well as innovative teaching-learning methods. The application of new teaching-learning methods such as problem-based teaching has proven to foster active student learning and to increase students’ self-regulation. These new methods do not only challenge students’ learning capacity but also teachers’ professional development and learning. Unlike traditional teaching methods which require teachers to explain the subject matter, innovative teaching-learning methods require teachers to fulfill different roles depending on the method applied. Thus, teacher learning and the investigation of factors contributing to teachers’ professional development are of high importance. Vermunt reviews factors which have been empirically shown to be related to teachers’ learning and discusses issues such as the development of adequate measures of teacher learning to be addressed by future research.

Richter, Kunter, Klusmann, Lüdke, and Baumert examine teachers’ uptake of formal and informal learning opportunities across the teaching career by use of data from 1939 German secondary teachers in 198 schools. Results show that formal learning opportunities, that is, in-service training, are used most frequently by mid-
career teachers, whereas informal learning opportunities exhibit distinct patterns across the teaching career. Specifically, the use of professional literature increases with teacher age, but teacher collaboration decreases. Teachers’ work engagement and professional responsibilities are hypothesized to predict changes over the career. These variables partly predict uptake of learning opportunities, but in contrast to the hypothesis, they do not fully explain the age-related differences observed.

The final two contributions to this volume are more specific in their respective topic. Bromme, Pieschl, and Stahl are concerned with student learning. More specifically, they investigate how epistemological beliefs of students, that is, students’ beliefs about the nature of knowledge, affect their learning. It is assumed that more sophisticated beliefs are associated with better alignment between task complexity and the learning process. Participants in their study are biology and humanity students who were given a list of tasks of different complexity. For each task, they completed a questionnaire. The results show that there is a reasonable alignment in that students’ answers to the questionnaire are related to task complexity in a meaningful way. Furthermore, alignment is linked to epistemological beliefs. For example, students with sophisticated beliefs judge the use of deep processing learning strategies more important across all tasks.

The paper of Klapproth and Schaltz is concerned with the predictive validity of academic and vocational-training achievement. The authors present a review of 52 studies in which the quality of predictors of academic achievements are investigated at the level of primary school, secondary school as well as achievements in universities and/or vocational training programs. A general finding is that achievements at each level can be predicted with sufficient accuracy by the achievements at the preceding school level. However, there are large differences in that shorter time periods of prediction, standardized achievement tests (as compared to unstandardized measures) and achievements in secondary school (as compared to other levels) have a higher accuracy of prediction.

We are grateful to all colleagues for their contribution to this volume. We hope that this volume will offer an important perspective on the domains of assessment, training, and learning in the professional development of teachers.

AFFILIATIONS

Sabine Krolak-Schwerdt  
University of Luxembourg

Sabine Glock  
University of Luxembourg

Matthias Böhmer  
University of Luxembourg
INTRODUCTION

Teacher judgments of student achievement have a considerable impact on students’ learning experiences and educational trajectories. Also, many instructional decisions are determined by teachers’ subjective judgments of their students’ achievement. The ability to accurately gauge student outcomes is therefore one of the key characteristics of a good teacher.

In the first part of this chapter, we provide a comprehensive description of our heuristic model of teacher judgment accuracy, which was first introduced in our meta-analysis on this issue (Südkamp, Kaiser, & Möller, 2012). In addition, we summarize the key findings of the meta-analysis. Studies included in the meta-analysis were limited to field studies. In the second part of this chapter, we introduce an experimental approach to the study of teacher judgment accuracy. In our own empirical research, we used the Simulated Classroom, which is a computer simulation of a classroom situation. Here, factors relevant to real classroom situations (e.g., student achievement, motivation, gender, subject, number of students, lesson length, and content covered) can be experimentally manipulated. Again, we provide key empirical findings on teacher judgment accuracy as well as on potential moderators. Finally, advantages and disadvantages of the field and experimental approach are discussed.

A MODEL OF TEACHER JUDGMENT ACCURACY

Given the important implications of teacher judgments (see Artelt, 2013; in this book), the question of their accuracy is critical. Accurate assessment of students’ performance is a necessary condition for teachers being able to adapt their instructional practices, to make fair placement decisions, and to support the development of an appropriate academic self-concept. In order to arrange possible influences on teacher judgment accuracy systematically, we propose a heuristic model of teacher judgment accuracy, which is displayed in Figure 1.

Teacher judgment accuracy is at the core of the model. It represents the correspondence between teachers’ judgments of students’ academic achievement and
students’ actual academic achievement measured by a standardized test. Usually, the correlation between the two is used as a measure of this correspondence. However, other indicators, such as the average difference between teacher judgments and students’ actual performance, can also be used.

A student’s test performance is the result he or she achieves on an academic achievement test. On the one hand, this result may depend on student characteristics such as prior knowledge, motivation, and intelligence. On the other hand, it may depend on test characteristics such as subject area, the specific tasks set, or task difficulty. In turn, a teacher’s judgment may depend on teacher characteristics such as professional expertise or stereotypes about students, and/or on judgment characteristics (e.g., whether the teacher is asked to judge a specific student competency, such as oral reading fluency, or to provide a global judgment of academic ability).

According to our model, teacher judgment accuracy is also influenced by the correspondence between judgment characteristics and test characteristics (dashed line). Potentially, the achievement test may measure a very specific academic ability (e.g., arithmetic skills), whereas the focus of the teachers’ judgment task is broader (e.g., rating students’ overall ability in mathematics), making it more difficult for teachers to make an accurate judgment. Another relationship that may influence teacher judgment accuracy is the correspondence between teacher characteristics and student characteristics (e.g., gender, ethnicity).

In the next section, we give a more detailed description of the model, reporting key empirical results for each aspect. We first summarize research findings on teacher judgment accuracy and then describe how teacher characteristics, judgment characteristics, student characteristics, and test characteristics influence teacher judgment accuracy.
Teacher Judgment Accuracy

Most research on teacher judgment accuracy examines the relationship between teachers’ judgments of students’ achievement and students’ actual performance on measures of achievement in various subject areas. Although, in most studies, academic achievement is measured by a standardized achievement test, some studies have used self-constructed, less standardized tests (e.g., curriculum-based measurement [CBM] procedures, see below).

The most commonly reported measure quantifying the correspondence between teachers’ judgments and students’ actual achievement is the correlation between the two. Overall, moderate to high correlations are reported (Begeny et al., 2008; Demaray & Elliot, 1998; Feinberg & Shapiro, 2003). For example, Feinberg and Shapiro (2009) reported correlations of .59 and .60 between teachers’ judgments and students’ decoding skills and reading comprehension, as measured by subtests of the Woodcock-Johnson-III-Test of Achievement. In the same study, a correlation of .64 was found between students’ oral reading fluency as measured by a CBM procedure and teachers’ predictions of oral reading fluency. In a review of 16 studies, Hoge and Coladarci (1989) found a median correlation of .66 between teachers’ judgments and students’ achievement on a standardized test. In our recent meta-analysis on teacher judgment accuracy (Südkamp et al., 2012), the overall mean effect size was found to be .63. On the one hand, these results may be interpreted as indicating that teachers’ judgments are quite accurate; on the other hand, their judgments are evidently far from perfect, and more than two thirds of the variance in teachers’ judgments cannot be explained by student performance. Additionally, the correlations found ranged substantially across studies, from .28 to .92 (Hoge & Coladarci, 1989) and from -.03 to .84 (Südkamp et al., 2012).

Methodological differences between studies need to be taken into account when considering the differences in results across studies. We therefore focus on how teacher judgment accuracy relates to judgment and test characteristics as well as to teacher and student characteristics, as discussed in detail below.

Several authors have noted the problems of relying solely on the correlation between teachers’ judgments and students’ performance on standardized achievement tests. For example, if teachers systematically perceive their students to be more or less competent than indicated by their performance on objective measures, their judgments may still be highly correlated with students’ performance (Eckert, Dunn, Codding, Begeny, & Kleinmann, 2006; Feinberg & Shapiro, 2003; Graney, 2008).

Indeed, the findings of studies using indicators other than correlations as measures of teacher judgment accuracy suggest that teacher judgments are rather inaccurate. Bates and Nettelbeck (2001) subtracted students’ reading accuracy and reading comprehension scores on a standardized achievement test from teachers’ predictions of these scores. Teachers generally overestimated the performance of the six- to eight-year-old students; inspection of the difference scores revealed that this held to a greater extent for low-achieving readers than for average- and high-achieving
students. In a study by Eckert et al. (2006), CBM material was used as an indicator of students’ mathematics and reading skills. Teachers were asked to estimate students’ reading and mathematics level (mastery, instructional, or frustrational). This judgment was compared with students’ actual reading and mathematics level as measured by the CBM material via percentage agreement. The results indicated that teachers overestimated students’ performance across most mathematics skills and on reading material that was at or below grade level. In line with this result, Begeny et al. (2008) found that teachers’ judgments of students with average to low oral reading fluency scores were rather inaccurate, and Feinberg and Shapiro (2003) reported that teachers generally overestimated the performance of low-achieving readers.

Nevertheless, studies have revealed large interindividual differences in teachers’ ability to judge student performance accurately (Helmke & Schrader, 1987). For example, Lorenz and Artelt (2009) reported moderate average correlations between teacher judgments and student performance in reading and mathematics for a sample of 127 teachers. The standard deviation for the mean of the correlations was .30 for reading and .39 for mathematics. Some teachers showed very high judgment accuracy; others, very low judgment accuracy. These findings raise the question of which individual teacher characteristics are related to teacher judgment accuracy.

**Teachers’ Judgments**

As shown in our model, teachers’ judgments are thought to depend on individual teacher characteristics and judgment characteristics. In the following, we first consider teacher characteristics that have been associated with teacher judgment accuracy and then discuss the influence of judgment characteristics on teacher judgment accuracy. A teacher’s characteristics are thought to influence her or his judgment at various stages of the judgment process (e.g., reception, perception, interpretation). Teacher characteristics such as job experience (Impara & Plake, 1998), beliefs (Shavelson & Stern, 1981), professional goals (Schrader & Helmke, 2001), and teaching philosophy (Hoge & Coladarci, 1989) have been associated with teachers’ judgment processes in the literature. Although the variability in the accuracy of teachers’ judgments is well documented (Hoge & Coladarci, 1989; Helmke & Schrader, 1987; Südkamp et al., 2012), empirical research has not yet pinpointed teacher characteristics that influence judgment accuracy.

**Teacher Characteristics**

*Job experience and exposure to students.* Impara and Plake (1998) expected teachers with more years of teaching experience to be experts in judging the item difficulties of a science test. However, they found no relationship between years of science teaching experience and the difference in actual and predicted item performance. The studies by Demaray and Elliott (1998) and Mul holland and
Berliner (1992) substantiated these findings. As already noted by Hoge and Coladarci (1989), research on teacher judgment accuracy provides few clues as to whether differences in teacher judgment accuracy are attributable to teacher characteristics. In our meta-analysis on teacher judgment accuracy, we considered the main teacher characteristics that have been associated with teacher judgment accuracy. We expected years of teaching experience and length of exposure to the students rated (i.e., how long teachers had taught those students) to be positively related to teacher judgment accuracy.

**Teachers’ gender, and age.** In addition, we probed for effects of teacher age and gender on teacher judgment accuracy. As teacher characteristics had only been examined in a small number of studies at the time we conducted our meta-analysis, we were not able to study their effects within our analyses.

Teachers’ judgments also depend on the characteristics of the judgment made. Usually, these characteristics reflect methodological decisions been made by the authors of a study. For example, some studies (e.g., Demaray & Elliott, 1998) ask teachers to directly judge students’ performance (e.g., to estimate the number of correct responses on an achievement test for each student), whereas others ask teachers to rate students’ overall academic ability indirectly on a rating scale. These differences are summarized under the label “judgment characteristics” in our model. It can be assumed that different judgment characteristics affect the correspondence between teachers’ judgments and students’ academic achievement.

**Judgment Characteristics**

**Direct versus indirect teacher judgments.** According to Hoge and Coladarci (1989), a distinction must be made between direct and indirect teacher judgments. In some studies, teachers are asked to assess students’ academic achievement on a standardized achievement test by estimating the number of items each student will solve correctly (Helmke & Schrader, 1987). This approach can be considered a direct rating. In other studies, teachers are asked to rate students’ performance in a certain subject on a Likert-type rating scale (e.g., a 5-point rating scale; DuPaul, Rapport, & Perriello, 1991). Hoge and Coladarci (1989) refer to this type of approach as indirect rating. In line with the results of Hoge and Coladarci, Feinberg and Shapiro (2003, 2009) and Demaray and Elliott (1998) found higher correlations for direct teacher judgments than for indirect teacher judgments. For example, Feinberg and Shapiro (2003) found a correlation of \( r = .70 \) between students’ test performance and direct teacher judgments, whereas the correlation with indirect teacher judgments was \( r = .62 \).

**Points on the rating scale.** Studies using rating scales to obtain teacher judgments differ in terms of the number of points on the rating scales implemented. Rating scales with many categories permit a sophisticated judgment, whereas scales with fewer categories allow a more global judgment. Generally, slightly higher
correlations with students’ actual performance are obtained for more sophisticated judgments than for more global judgments. To date, this variable has been neglected in empirical research on teacher judgment accuracy.

**Judgment specificity.** According to Hoge and Coladarci (1989), the distinction between direct and indirect teacher judgments also has implications for the specificity of the judgment. In general, direct judgments are more specific than are indirect judgments, as they are explicitly tied to the criterion in the judgment process. Indirect teacher judgments may be differentiated in terms of their degree of specificity. Following the approach used by Hoge and Coladarci, teachers’ judgments can be allocated to one of five categories, ranging from low to high specificity. First, a judgment that requires teachers to rate students’ academic achievement on a rating scale (e.g., poor – excellent) is considered to be of low specificity. Second, in a ranking, the teacher’s task is to put the students of his or her class into rank order according to their achievement. Third, tasks requiring teachers to find grade equivalents for students’ performance on a standardized achievement test are considered to be of average specificity. Fourth, tasks asking teachers to estimate the number of correct responses achieved by a student on a standardized achievement test are slightly less specific than the fifth and most specific category, in which teachers indicate students’ item responses on each item of an achievement test. In their review, Hoge and Coladarci found a median correlation of .61 for studies using ratings, which was the predominant approach. The median correlations for studies using rank ordering (median r = .76), grade equivalents (median r = .70), number of correct responses (r = .67, for a single study), and item-based judgments (median r = .70) were indeed higher.

**Norm-referenced vs. peer-independent judgments.** In addition, teacher judgments may differ in whether they are norm-referenced or peer-independent. For example, Helwig et al. (2001) asked teachers to rate students’ academic achievement on an absolute scale (very low proficiency – very high proficiency), whereas Hecht and Greenfield (2002) asked teachers to estimate students’ academic achievement in relation to other members of the class (in the bottom 10% of the class – in the top 10% of the class). Hoge and Coladarci (1989) considered this aspect in their meta-analysis, but found no substantial difference between correlations. The median correlation for norm-referenced judgments was .68; that for peer-independent judgments was .64.

**Domain specificity.** Finally, teacher judgments differ in terms of their domain specificity. Whereas some studies ask teachers to judge students on a very specific ability (e.g., arithmetic skills; Karing, 2009), others ask them to judge students’ overall academic achievement (e.g., Li, Pfeiffer, Petscher, Kumtepe, & Mo, 2008). To our knowledge, no studies to date have examined the influence of the domain specificity of teachers’ judgments on teacher judgment accuracy. However, it seems
reasonable to hypothesize that it is easier to make a focused judgment on a domain-specific ability than to judge a student’s overall academic ability. Instead of distinguishing between direct and indirect teacher judgments, we used a slightly different categorization in our own meta-analysis. We distinguished between informed versus uninformed teacher judgments, considering whether teachers were informed or uninformed about the content of the test prior to rating students’ academic achievement and found significantly higher correlations between teachers’ judgments and students’ test performance for informed than for uninformed teacher judgments. No effects of the other judgment characteristics (i.e., number of points on rating scales, judgment specificity, norm-referenced versus peer-independent judgments) were found.

Students’ Test Performance

Teacher judgment accuracy also depends on students’ test performance, which in turn depends on individual student characteristics and characteristics of the test.

Student Characteristics

Several student characteristics have been identified as influencing the accuracy of teachers’ judgments. For example, Bennett, Gottesman, Rock, and Cerullo (1993) found that teachers who perceived their students to exhibit bad behavior also perceived these students as low academic performers, regardless of the students’ academic skills. In a study by Hurwitz, Elliott, and Braden (2007), the accuracy of teachers’ judgments was related to students’ disability status: Teachers predicted the mathematics test performance of students without disabilities more accurately than that of students with disabilities. Ritts, Patterson, and Tubbs (1992) found that teachers also took student attractiveness into account when judging students’ academic achievement.

Gender. Although many studies have examined student gender as a potential moderator of teacher judgments, no consistent significant effects of student gender on teachers’ judgments have been established (Demaray & Elliott, 1998; Hoge & Butcher, 1984). A large body of research in this area focuses on teacher expectations rather than teacher judgment accuracy. However, there is considerable overlap between the methodologies used. In both types of studies, teachers are asked to judge students’ academic performance. Whereas teacher expectancy research tends to focus on differences in teachers’ judgments of different groups (e.g., European vs. African American students, girls vs. boys), research on teacher judgment accuracy puts less emphasis on distinguishing between groups of students on the basis of a certain characteristic, focusing rather on the correspondence of teacher judgments with students’ actual academic achievement. In teacher expectancy research, differences in teacher judgments of different groups are sometimes, but not
always, controlled for students’ actual academic achievement. Nevertheless, teacher expectancy research holds important implications for our meta-analysis, as it reveals factors influencing teachers’ judgment. In their meta-analysis of teacher expectancy effects, Dusek and Joseph (1983) reviewed 16 studies in which student gender was related to a measure of teacher expectations of students’ academic performance. The results indicated that student gender does not affect teacher expectancies of general academic achievement. Jussim and Eccles (1995) reported that teachers perceived girls in fifth grade mathematics classes as performing slightly higher than boys, but this perception was accurate, as girls in their study slightly outperformed boys on standardized achievement tests. In contrast, Tiedemann (2000) reported that teachers rated the mathematical ability of third and fourth grade boys to be slightly higher than that of their female counterparts, although these girls’ and boys’ grades in the previous school year were not significantly different. Controlling for first grade students’ performance on a standardized reading test, Hinnant, O’Brien, and Ghazarian (2009) found that teachers tended to overestimate the reading ability of girls and to underestimate that of boys. In sum, there is only little evidence that teacher judgment accuracy is influenced by student gender.

Grade level. Studies have also examined whether teachers’ judgment accuracy differs across grades. Kenny and Chekaluk (1993) reported that teachers’ assignment of second grade children to reading categories (advanced, average, poor) was more accurate than their assignment of first graders and kindergarteners. In contrast, Maguin and Loeber (1996) reported significantly higher correlations between teacher judgments and students’ reading and mathematics achievement for first graders than for fourth and seventh graders. Kuklinski and Weinstein (2001) reported that teacher expectancies accentuated achievement differences to a greater extent in the early elementary grades than in the later elementary grades.

Ethnicity. Research has shown that teacher expectancies of students are influenced by students’ ethnicity (Baron, Tom, & Cooper, 1985; Dusek & Joseph, 1983). A recent meta-analysis of studies conducted in the United States (Tenenbaum & Ruck, 2007) revealed that teachers had more positive expectations for European American children than for ethnic minority children. In turn, Chang and Sue (2003) found that teachers had more positive expectations for Asian American students than for other ethnic groups. Wigfield, Galper, Denton, and Seefeldt (1999) showed that teachers’ beliefs about students’ academic abilities significantly predicted students’ performance on a standardized test, and that teachers’ ratings differed significantly between ethnic groups, with European American children being rated significantly higher than Hispanic children or African American children.

Socioeconomic background. The literature on teacher expectancies has identified students’ social background as a factor that crucially informs teacher expectancies (Alexander, Entwisle, & Thompson, 1987; Jussim, Eccles, & Madon, 1996).
Alvidrez and Weinstein (1999) reported that teachers judged children from higher socioeconomic backgrounds more positively and students from lower socioeconomic backgrounds more negatively than the students’ performance on the Wechsler Intelligence Scales would predict. Kennedy (1995) found the proportion of low-income students at a school to be strongly negatively correlated with teachers’ perceptions of students’ ability. In contrast to their expectations, Wigfield et al. (1999) did not find teachers’ judgments to differ between former Head Start children from socioeconomically disadvantaged families and non-Head Start children. Likewise, Jussim and Eccles (1995) found no evidence that teachers judged students with lower socioeconomic status any less favorably than students with higher socioeconomic status. Hinnant et al. (2009) argued that teachers’ expectations may influence the reading performance of minority groups in particular. In their study, first grade teachers’ expectations were reliably linked to the third grade performance of minority boys, but not of White students or Non-White girls. In mathematics, teacher expectations were significantly and positively related to the later mathematics performance of children from families with low or average income, but unrelated to that of high-income children. In a study with a sample of kindergarten children from various ethnic groups living in low-income families, Hauser-Cram, Sirin, and Stipek (2003) found that teachers rated the children as less competent if they perceived value differences between themselves and the parents.

As is the case for teacher characteristics, however, few studies to date have reported information on the student sample. Moreover, any data available are not readily comparable across studies (e.g., only the percentage of female/male students was reported). Therefore, we decided not to conduct moderator analyses on student characteristics in this meta-analysis.

Test Characteristics

As shown in our model, students’ test performance also depends on characteristics of the test. Like the judgment characteristics summarized above, these test characteristics in turn depend on methodological decisions made by the author(s) of the studies. In studies on teacher judgment accuracy, various instruments are used to measure students’ academic achievement, ranging from highly specific tests measuring, for example, receptive vocabulary (e.g., the Peabody Picture Vocabulary Test used by Fletcher, Tannock, & Bishop, 2001) to broader tests measuring students’ performance in different subject areas (e.g., the Kaufman Test of Academic Achievement measuring achievement in mathematics, reading, and spelling used by Demaray & Elliott, 1998). Our model summarizes such differences between tests under the label “test characteristics.” Various test characteristics can be assumed to influence the correspondence between teachers’ judgments and students’ performance.

Subject matter Comparing correlations between teachers’ judgments and students’ academic achievement in different subjects, Hopkins, George, and Williams (1985)
found that correlations were significantly lower for social studies and science than for language arts, reading, and mathematics. Using CBM procedures to gauge students’ academic achievement, Eckert et al. (2006) found higher correlations for reading than for mathematics. In turn, Coladarci (1986) reported teachers’ judgments to be more accurate for students’ performance in mathematics computations than for mathematics concept items. Demaray and Elliott (1998) found no difference between correlations in language arts and in mathematics. Hinnant et al. (2009) found that teachers’ ratings of academic ability as measured by an academic skills questionnaire were highly correlated with standardized measures of achievement in reading ($r = .53–.67$) and mathematics ($r = .54–.57$). Evidently, the empirical findings on the influence of subject matter on teacher judgment accuracy are inconsistent. In addition, there are very few studies focusing on subjects other than language arts and mathematics. As exceptions, achievement in sports (swimming) was measured in a study by Trouilloud, Sarrazin, Martinek, & Guillet, (2002) yielding a comparably high correlation between teachers’ judgments and students achievement in swimming ($r = .78$). Music achievement was measured in a study by Klinedinst (1991), which reported a comparably low correlation ($r = .21$).

**CBM procedures vs. standardized achievement tests.** Some studies of teacher judgment accuracy have used CBM procedures as indicators of students’ achievement (Eckert et al., 2006; Feinberg & Shapiro, 2003; Hamilton & Shinn, 2003). According to Feinberg and Shapiro (2003), CBM is closely linked to actual in-class student performance, as methods derived from curriculum materials provide a closer overlap with the content of instruction than do published norm-referenced tests. For example, Feinberg and Shapiro (2009) used three reading probes of 150 words to measure students’ oral reading fluency. The number of words read correctly per minute was used as a measure of students’ reading performance. Mispronunciations, omissions, and substitutions were counted as errors. The median for each of the three probes was computed and used as the student’s overall CBM score. The authors argued that teachers are likely to use students’ observed classroom behavior as their basis for judging students’ academic achievement. They criticized the “lack of a potential overlap between content assessed on a standardized test and student behavior” (Feinberg & Shapiro, 2003; p. 54), which complicates accurate teacher judgments. Indeed, students’ achievement on curriculum-based testing material has been compared with their achievement on published, standardized achievement tests. Feinberg and Shapiro (2009) found that correlations between a CBM procedure measuring oral reading fluency and teachers’ predictions of oral reading fluency were slightly higher ($r = .64$) than correlations between a global teacher rating of students’ performance and two subtests of a standardized achievement test ($r = .59$ and $r = .60$).

**Domain specificity.** Like teacher judgments, academic achievement tests also differ in terms of their domain specificity. Whereas some tests are designed to measure a
very specific academic ability (e.g., phonological awareness; Bailey & Drummond, 2006), others measure different aspects of academic ability (e.g., the Woodcock-Johnson Achievement Battery; Benner & Mistry, 2007).

In terms of test characteristics, we found no evidence for a difference in teacher judgment accuracy between language arts and mathematics in our meta-analysis. The effects of the other test characteristics were not significant either. Therefore, results are generalizable across several types of judgments and tests.

**Correspondence Between Judgment and Test Characteristics**

As depicted in our model, the correspondence between judgment characteristics and test characteristics is assumed to influence teacher judgment accuracy.

*Time gap.* In their review, Hoge and Coladarci (1989) included only studies in which the achievement test was administered at the same time as the teacher rating task. In many studies, however, these two measures are not implemented concurrently (Pomplun, 2004). Due to temporal proximity, we expect to find higher correlations between teachers' judgments and students' academic achievement when both measures are administered concurrently than when the test is administered either before or after the rating task.

*Congruence in domain specificity.* Finally, we considered the congruence in the domain specificity of the teacher rating task and the achievement test. Theoretically, the achievement test may measure a specific academic ability whereas the teacher judgment task may be less specific—or vice versa. For example, Hecht and Greenfield (2001) found teachers' judgments of students' overall academic competence to be correlated with the students' performance on the Letter-Word-Identification subtest of the Woodcock-Johnson Test of Achievement-Revised. Here, a general judgment is set in relation to a very specific ability. We expected to find higher correlations between teachers' judgments and students' achievement in studies in which the domain specificity of the teacher rating task and the achievement test was congruent (e.g., teachers rated students' reading comprehension; students were administered a test of reading comprehension), and lower correlations in studies in which the domain specificity was incongruent (e.g., teachers rated students' overall academic achievement; students were administered a test of reading comprehension).

As expected, the congruence between the teachers' rating task and the achievement test administered to students was related to teacher judgment accuracy, with higher congruence being associated with higher accuracy levels. Because the match between teachers' judgments and students' test performance was higher when both measures addressed the same domain and same ability within a domain, it is reasonable to assume that a "mismatch" leads to lower teacher judgment accuracy.
Correspondence Between Teacher Characteristics and Student Characteristics

To complete the picture of teacher judgment accuracy, our model includes the correspondence between the teacher characteristics and the student characteristics influencing teacher judgment accuracy. To our knowledge, few studies have taken this aspect into account, and their results have been mixed. In their study of teachers’ expectations of the reading and mathematics achievement of their students, Alexander et al. (1987) took teacher and student race (Black vs. White) and socioeconomic status into account. The authors found that low-status and Black students were evaluated less favorably than high-status and White students, especially by high-status teachers. Chang and Sue (2003) studied the effects of student race on teachers’ assessment of student behavior using vignettes that were paired with a photograph of a child. Due to unequal sample sizes (74.1% European Americans, 10.6% Hispanic, 3.6% African American), the authors were unable to analyze the ratings with respect to teacher ethnicity. In another study, Chang and Demyan (2007) overcame this problem by dichotomizing teachers’ race into “White” and “ethnic minority.” Teachers rated the students on personal traits; no significant interactions between teacher race and child race were found for any of the 15 traits.

Unfortunately, we were not able to study the effects of the correspondence between teacher characteristics and student characteristics in our meta-analysis as comparable data (e.g., on ethnicity) were not reported in a sufficient number of studies.

EXPERIMENTAL RESEARCH ON TEACHER JUDGMENT ACCURACY

Besides the field studies introduced above, teacher judgment accuracy has also been studied in experimental studies, which allows for inspecting teachers’ information processing and decision making more closely (e.g., Krolak-Schwerdt, Böhmer, & Gräsel, 2013; Südkamp & Möller, 2009). In our own empirical research, we have been using a simulated classroom paradigm (Südkamp, Möller, & Pohlmann, 2008).

The Simulated Classroom

The Simulated Classroom (see also Fiedler, Freytag, & Unkelbach, 2007; Fiedler, Walthier, Freytag, & Plessner, 2002) is a computer simulation of an instructional situation in which student factors—e.g., achievement (in terms of the proportion of correct answers), motivation (in terms of participation in class), gender (as indicated by a photograph or name)—and instructional factors (subject, number of students, lesson length, content covered) can be experimentally manipulated. The Simulated Classroom is programmed in Java; participants work individually on personal computers. To begin, participants are given an introduction to the functioning of the Simulated Classroom, in which they take on the role of a teacher. Before the lesson starts, they are informed about the students’ grade and the topic of the lesson. Their
task is to select questions on that topic from a menu of possible questions and to address these questions to the students in their “class.” The students are represented by names on virtual desks (see Fig. 2).

![Screenshot of the Simulated Classroom.](image)

The names were chosen at random from the most popular children’s names in Germany in the year the simulated student would theoretically have been born. Photographs, names (and thus gender), and seating positions were allocated at random (making sure that the gender and name allocated matched the photo). Each question selected by the “teacher” is displayed at the bottom left of the computer screen. Tasks were either taken from standardized achievement tests or developed specifically according to the study’s requirements. In our studies, we mainly focused on the subjects of mathematics and language arts. Bringing together experts in the field of chemistry didactics and educational psychology, Bolte, Köppen, Möller, and Südkamp (2011) developed the Simulated Science Classroom, which incorporates tasks on specific science concepts (e.g. the particle model of matter or concepts of inquiry). When a question has been selected, students volunteer to answer the question in accordance with their predefined motivation parameters. These students are indicated by the coloring of their desks, which changes to yellow. Depending on the study’s aim the “teacher” may call on any of the students (whether or not they have volunteered an answer) or just on students volunteering to answer the question, by a mouse click on the respective desk. That student then gives a correct or an incorrect answer depending on his or her predefined ability parameter. The answer is displayed at the bottom right
of the screen. If it is correct, it appears in a green box (see Fig. 2). Otherwise, one of several possible incorrect answers appears in a red box. This variation of incorrect answers reduces the probability that the same incorrect answer will be given consecutively by different students. Once a question-and-answer sequence has been completed, the “teacher” can direct either the same question or a new question to any of the students. The length of the lesson can be varied; the teacher can ask any number of questions.

The proportion of correct answers provided by each student is experimentally varied and represents the level of student achievement (e.g., probability of a correct answer approximately .80 for high achievement). The simulated students’ achievement behavior in the Simulated Classroom is determined by a probability algorithm, such that the proportion of correct or incorrect answers given by each student corresponded approximately with his or her achievement parameter. In most of our studies, we also varied student motivation experimentally, which is operationalized as the probability of a simulated student volunteering to answer a question (e.g., probability of volunteering .20 for low motivation). Motivation behavior in the Simulated Classroom is again determined by a probability algorithm, such that proportion of questions each student volunteered to answer corresponded approximately with his or her motivation parameter. Participants usually have 16-18 minutes to get a picture of the simulated students’ academic achievement and motivation.

At the end of the “lesson,” participants are asked to judge the proportion of correct answers given by each student. Ratings are usually given on a scale from 0 to 100%. It is thus possible to gauge the extent to which the participants’ judgments correspond with the students’ actual achievement. Likewise, the proportion of questions that each student volunteers to answer serves as a measure of student motivation. Participants are also asked to judge this proportion, thus allowing perceived and actual student motivation to be compared. Ratings are again given on a percentage scale. To get an indirect judgment, teachers are also asked to grade the simulated students.

Teacher Judgment Accuracy within the Simulated Classroom

Our findings on the accuracy of judgments of student achievement indicate that teachers and teacher candidates are fairly successful in gauging individual students’ relative achievement level in the Simulated Classroom. The correlations between actual student achievement and judgments of achievement are usually moderate to high in size. For example, Südkamp et al. (2008) found correlations of .62 (first simulated lesson) and .68 (second simulated lesson), while Kaiser, Retelsdorf, Südkamp, and Möller (2013) found correlations of .69 (second partial study) and .57 (third partial study). These finding are consistent with the results of field studies on the accuracy of teacher judgments of student achievement (Hoge & Coladarci, 1989; Südkamp et al., 2012).
Moderators of Teacher Judgment Accuracy

Concerning moderators of teachers’ judgment accuracy, we will focus on the influence of student characteristics and teacher characteristics here, as those factors could not be included in our analyses within the meta-analysis of field studies. *Student characteristics: Motivation and ethnicity.* In a recent study based on the simulated classroom paradigm (Kaiser et al., 2013), we combined the field and experimental approach to the study of teacher judgment accuracy and examined whether students’ achievement influences teachers’ judgments of their motivation and vice versa. In the field study teachers and their students, who were tested and surveyed within a comprehensive project on the development of reading literacy and reading motivation at secondary level in Germany (e.g., Möller, Retelsdorf, Köller, & Marsh, 2011; Retelsdorf, Becker, Köller, & Möller, 2011; Retelsdorf, Köller, & Möller, 2011). Measures of students’ academic achievement and motivation were implemented, while teachers’ judged students’ academic achievement and motivation respectively. In two experimental studies, teacher candidates worked on the Simulated Classroom. Here, the academic achievement and motivation of simulated students was varied experimentally. In all three studies, structural equation modeling revealed an effect of student achievement on teachers’ judgments of student motivation and an effect of student motivation on teacher judgments of student achievement—above and beyond the association of each student characteristic with teachers’ judgments of that characteristic.

In two studies on the influence of students’ ethnicity on teachers’ judgments of students’ academic achievement, we varied students’ ethnicity within the Simulated Classroom (Kaiser, Schubert, Südkamp, & Möller, 2012). In the first study, there were eight German students, one Asian student, and three Turkish students within the Simulated Classroom, which was indicated by their pictures and traditional names. In Germany, having a Turkish migration background is still associated with lower academic achievement in comparison to non-migrant students and with lower academic opportunities. According to this stereotype, we expected Turkish students to be judged less positively compared to their German classmates, while controlling for the students’ achievement within the Simulated Classroom. Results showed that low achieving Turkish students were judged less favorably than low achieving German students, but also more accurately. In contrast, high achieving Turkish students were judged more positively than high achieving German students, meaning Turkish students were judged more accurately.

The second study was conducted analogous to the first study. Here, there were eight German students, one Turkish student and three Asian students in the Simulated Classroom. In western society, Asians are often perceived to be a “model minority” showing high effort and achievement in academics. According to this stereotype, we expected the Asian students within the classroom to be judged more positively than their German classmates. Results showed that low achieving Asian students were judged less favorably than low achieving German students, but also more accurately.
On the other hand, high achieving Asian students were judged more positively than high achieving German students, again meaning Asian students were judged more accurately. As we didn’t find the expected general positive bias in judgments towards Asian students, we also analyzed whether it could be that being a minority is to account for higher judgment accuracy. So we estimated a path model with manifest variables using the Mplus software (Muthén & Muthén, 2010). We analyzed whether being a minority moderated the relationship between students’ actual achievement and teachers’ judgment of student achievement. The results revealed a significant path from students’ achievement to teachers’ judgments, no significant path from the students’ minority status, but a significant interaction-path. The significant path from students’ actual achievement to teachers’ judgments shows, that teachers were quite accurate in judging students’ achievement. But the significant interaction reveals that the relationship between students’ actual achievement and teachers’ judgment is more accurate for minority students.

Teacher characteristics: Cognitive abilities. As mentioned above, there are only a few studies, which explicitly focus on teacher characteristics as moderators of teacher judgment accuracy. To our knowledge, teachers’ cognitive abilities as a prerequisite of judgment accuracy have not been a focus of research at all. Since teachers are confronted with much diagnostic information during their lessons, a high information processing speed could be a necessary condition to be able to judge students’ characteristics accurately. The aim of this study (see Kaiser, Helm, Retelsdorf, Südkamp, & Möller, 2012 for a more detailed description) was to give a first hint on the relationship between teachers’ diagnostic and cognitive abilities. The diagnostic ability of teacher students at the University of Kiel was tested within the Simulated Classroom paradigm. Preceding the Simulated Classroom the subjects’ cognitive abilities were measured with 34 selected items of the Advanced Progressive Matrices (APM; Raven, 1962) for one sample. In another sample comprising teacher students the cognitive abilities were measured with the subscale Figure Analogies of the German version of the Cognitive Abilities Test (Heller & Perleth, 2000). In order to test the relation between cognitive and diagnostic abilities a multi level-analysis was conducted. On level 1 students’ actual achievement was used to predict teachers’ judgments (as a measure of judgment accuracy). On level 2 the cognitive abilities were taken into account and it was tested whether there is a relation between cognitive abilities and judgment accuracy. The multi level-analysis revealed a significant cross-level-interaction for both samples showing that teachers’ higher cognitive abilities were associated with higher judgment accuracy.

SUMMARY AND OUTLOOK

In this book chapter, we first extended the description of our heuristic model of teacher judgment accuracy, which was briefly introduced in the discussion of Südkamp et al. (2012). In line with the model, we brought together findings on teacher judgment
accuracy from field studies. As empirical research findings of 75 studies on the issue are statistically summarized in our meta-analysis, we highlight the most important findings here. Second, we introduced an experimental approach to the study of teacher judgment accuracy. In our own empirical research, the Simulated Classroom proved to be a useful instrument for examining teacher judgment accuracy. The instrument makes it possible to experimentally manipulate student characteristics, and thus complements studies conducted in real-life contexts: Psychological phenomena that have been observed in field studies to be associated with teacher judgments of students can be investigated more closely under experimental conditions (Wang, Treat, & Brownell, 2008).

Although our findings concerning teacher judgment accuracy are similar in field and experimental studies, it remains questionable to what extent the same findings can be expected in real-life classrooms as in the Simulated Classroom. In the following, we discuss some differences between teacher judgments in the two contexts. Whereas in real life, teachers’ judgments are based on a wealth of informal observations, in the Simulated Classroom, it is possible to experimentally control the information on which these judgments are based. The observations made in the Simulated Classroom are thus far less complex than those made in real-life classroom situations, in which the demands on teachers in terms of the collection and interpretation of diagnostic information are much higher. The reduced complexity of the Simulated Classroom may thus be a further explanation for the slightly higher levels of judgment accuracy observed in the experimental studies. Moreover, the systematic distribution of student achievement and motivation implemented in the Simulated Classroom does not reflect that found in a natural environment. The instrument can, however, be extended to allow the targeted variation of various student and class characteristics, thus allowing a differentiated analysis of information processing processes in teachers and teacher candidates. The Simulated Classroom is therefore considered a promising tool for the research concerning teachers’ judgment accuracy (Schrader, 2010; Spinath, 2012).

The research presented in this book chapter brings together different research strategies. However, the question whether the same construct is measured with the different approaches still needs to be clarified empirically. This will be the focus of future projects. As a long-term objective the Simulated Classroom could be used as a training instrument. Teachers and teacher candidates could get immediate feedback regarding their judgment accuracy and as a result decrease erroneous judgment tendencies.

REFERENCES/BIBLIOGRAPHY


TEACHERS’ JUDGMENTS OF STUDENTS’ ACADEMIC ACHIEVEMENT


AFFILIATIONS

Anna Südkamp
Rehabilitation Psychology
TU Dortmund University, Germany

Johanna Kaiser
Educational Psychology
University of Kiel, Germany

Jens Möller
Educational Psychology
University of Kiel, Germany