

# Teacher Judgments and Their Role in the Educational Process

CORDULA ARTELT

## Abstract

Teacher judgments matter for students: They have an impact not only on students' academic self-concepts and for their theories about themselves as learners but also on educational pathways through grades and school leaving certificates. However, do teachers have the expertise for such judgments? Is there such a thing as a professional ability to judge (diagnostic competence)? The essay provides an overview on research related to teacher judgment accuracy and bias and specifies conditions under which accurate teacher judgments can (and cannot) be expected. It is argued that in research as well as in teacher training, causes of judgment and judgment demands need to be taken into account and that the role of knowledge components in controlled and automatic decision-making after intensive experience on the job needs to be better understood.

## PREVALENCE AND RELEVANCE OF TEACHER JUDGMENTS

Teacher judgments are omnipresent in everyday school life. During lessons, on average, teachers ask a question or set a task every 30 s (Niegemann & Stadler, 2001). The more adaptive these actions are, the more (implicit or explicit) judgments teachers have to make of students' achievement level, their learning trajectories, and the related difficulty of specific tasks and learning materials. However, not only adaptive approaches set high standards for past and/or concurrent decision-making: All forms of teachers' task setting, question generating, and feedback on students' performance rely on assumptions about students' (age-related) performance and abilities, the expected learning gains of specific treatments, and the assumed difficulty level of certain tasks or questions.<sup>1</sup> These tasks thus imply a form of teachers' decision-making. Some of these judgments are related

1. These expectations might not be the result of individual teachers' decision-making but instead be part of their professional knowledge (e.g., about typical achievements of a K4 student in mathematics) gained during teacher training or based on curriculum/educational standards and textbooks.

primarily to educational goals (e.g., initiating assessment aimed at optimizing students' learning) and are labeled *formative assessment*, sometimes also *assessment for learning* (Looney, 2011; OECD, 2012). Together with these teacher judgments and decisions related to the improvement of the quality of teaching (e.g., adaptive teaching) and to students' learning processes, the evaluation of individual students' performance is a core component of teachers' professional actions and duties (*assessment of learning*). This primarily societal function of student evaluations also leads to selection or placement decisions (e.g., concerning school career) and finds its expression in granting certificates (summative assessment). Especially in countries with a differentiated secondary school system, assignment to school types or school tracks depends largely on grades, school leaving certificates, and teacher recommendations. Moreover, selection decision for specific programs or needs (special educational needs, giftedness) also depends in part on teacher judgments.

#### INFLUENCES OF TEACHER JUDGMENTS ON STUDENTS

Teacher judgments matter for students. They influence not only students' educational opportunities but also their selves and attitudes.

How do teachers shape educational trajectories? Teachers directly impact on their students' educational careers via formal assessments and school leaving grades, which in turn enable students to select specific academic tracks. In addition, teachers indirectly influence students' educational trajectories via (more or less formal) recommendations to parents related to their children's school choices. In most cases, it is likely that parents trust teacher recommendations and act accordingly. Teachers thus play a crucial role during educational transition phases. For that reason, the stronger the influence of features other than those relevant for the decision (i.e., social categories, stereotypes; see the following paragraphs) on teachers' decision-making, the more frequently the educational opportunities are displaced.

Studies of the role of teacher judgments also need to take psychological effects into account. The way in which teachers judge and evaluate individual students' performance and behavior impacts on students' selves. These evaluations and interpretations of students' behavior and performance are fundamental in shaping their academic self-concepts and self-esteem. This is partly due to the feedback students receive on their performance, such as information about the relative position within a class (social reference norm) or the absolute level of their performance (e.g., 80% correct). However, on top of this frequently described effect of achievement-related feedback on students' academic self-concepts, the way in which this information is framed, communicated and further used by the teacher in classes has

an additional influence on students' motivation. Why so? In their every-day practice, teachers build on expectations about their students. These expectations (e.g., related to future performance in a specific domain) are, among others, related to teachers' epistemological beliefs (Hofer, 2001) and to their causal attributions<sup>2</sup> to students' achievement-related success or failure (Stiensmeier-Pelster & Heckhausen, 2008; Zhou & Urhahne, 2012). Teachers who incorrectly interpret a student's failure as a lack of ability (internal/stable) or effort (internal/variable) are likely to be biased not only when allocating grades but also in terms of further achievement expectations. Expectancy-confirmation processes on the part of the teacher (Darley & Fazio, 1980) might thus lead to a chain of reactions and overt behavior toward students that affects the latter's own self-evaluations and thus their motivation. Teachers seem to form blueprints for students' theories about their own capabilities. In that sense, all achievement-related judgments of teachers are, psychologically speaking, high stake.

#### IS THERE SUCH A THING AS TEACHERS' DIAGNOSTIC COMPETENCE (ASSESSMENT LITERACY?)

The notion of teacher judgment accuracy having an impact on students is based on the idea that judgment accuracy is not a random variable that varies according to situation, assessments demands, and other characteristics of the teachers and students, but rather a disposition or trait (at least in certain domains). Hence, in the tradition of teacher professionalism—especially in the German literature—the term *diagnostic competence* is frequently used (Baumert & Kunter, 2006). Diagnostic competence (or assessment literacy) is regarded as a key competence in the context of teaching and learning (Weinert, Schrader, & Helmke, 1990). In order to meet different students' needs, provide accurate feedback, and tailor instructional designs accordingly, teachers need to be able to correctly judge both relevant student characteristics and the difficulty level and demands of class material. Teachers with high accuracy in these judgments are therefore presumed to foster students' learning processes indirectly. Such effects of teacher judgment accuracy on student learning gains have also been reported in the literature (Karing, Pfof, & Artelt, 2013).

In recent years, however, a lively debate has ensued about the concept of diagnostic competence. One of the reasons why it has been criticized is the lack of empirical support for the assumption of a homogeneous competence dimension. It appears that teacher judgment accuracy varies as a function

---

2. This itself is a form of (teacher) judgment.

of the respective achievement/competence domain under study (e.g., students' mathematical or reading competence and their motivation, anxiety, or intelligence; e.g., Spinath, 2005) and—to some degree—of other features and demands of the judgment process (see the following section). To this end, there is reason to believe that *diagnostic competence* is not an adequate term for these intraindividually varying judgment qualities. Whether it is reasonable to use the term *diagnostic competence* (as a trait or disposition) or to use the term *judgment accuracy* instead (as a state) thus remains a matter of debate. At the same time, empirical evidence concerning teachers' diagnostic competence is incomplete and sometimes contradictory. As is argued throughout this essay, these mixed results can be partly attributed to the fact that judgment purposes and demands are not fully taken into account in the current literature. Teachers should not be expected to deliver high quality and accurate judgments independent of the goals, demands, and standards placed on their judgments. In particular, the research aimed at assessing teachers' diagnostic competence does not seem to take this sufficiently into account. Moreover, research on the quality and accuracy of teacher judgments does not yet have adequate standards for evaluating accuracy.

#### ASSESSING THE ACCURACY OF TEACHER JUDGMENTS

When judging whether Mary's answer (60) to her teacher's question ( $34 + 29 - 2$ ) is correct, the answer is fairly obvious. However, whether the teacher was correct in grading Mary's mathematical competence and assigning her to a specific program is far more difficult to determine. Obviously, a criterion is needed for this evaluation.

A common approach in the literature is to evaluate judgment accuracy by comparing teacher judgments related to student achievement with students' achievement scores in a standardized test in the respective domain (test as criterion). The use of standardized tests has many advantages. However, it should be kept in mind that the validity of the standard depends on the kind of teacher judgment. In times of increasing assessments of educational effectiveness, accountability, and the relevance of standardized testing in this regard, it is important to keep in mind that objective and standardized tests do not guarantee a valid benchmark for teacher judgments per se. Such tests vary not only in scope and quality but also with respect to the degree to which they can serve as a valid criterion for students' characteristics (i.e., competencies/abilities).

Another approach for assessing the validity of teacher judgments uses social disparities after institutional transitions as criterion. The degree to which the children's attendance of different school tracks varies as a function of social status, even after controlling for children's grades and

school achievement (which should be the only criterion for the assignment), is often taken as an indicator of social disparities. To some extent, these are due to disparities in teachers' grading practice and their formation of educational recommendations. On the basis of this criterion, research conducted in highly tracked school systems such as Germany has shown that teacher judgments are in fact influenced by students' social origin and that chances of attending a Gymnasium (academic school track) vary accordingly (Baumert, 2016).

Research using students' results on a standardized test as a criterion for teacher judgments makes use of different indicators of judgment accuracy. The most common indicator measures rank-order accuracy. Here, teacher ratings on students' individual (or group level) abilities are correlated with students' individual performance on a standardized test; a higher correlation indicates higher judgment accuracy, indicating the degree to which the correct rank order is reproduced. Because this indicator is predominantly used, the following passage focuses on findings based on it. However, although a teacher might perfectly well reproduce the (achievement-related) rank order, he or she might perform poorly when judging the level of performance (e.g., number of items solved) or when judging the variation of scores within a class. A more complete picture of the accuracy of teacher judgments thus takes into account the rank order, the level component, and the variation component as largely independent indicators.

The majority of studies indicate that teachers' rank-order judgments of student characteristics correlate moderately highly with the independently measured rank order of their students: On the basis of their meta-analysis, Hoge and Coladarci (1989) reported correlations between  $r=0.28$  and  $r=0.92$ , with a median correlation of  $r=0.66$ . In a recent meta-analysis (Südkamp, Kaiser, & Möller, 2012), the Fisher's  $z$ -transformed correlations ranged between  $r=-0.03$  and  $r=1.18$ , with a median correlation of  $r=0.53$ . However, as can be seen by the range of results reported in the two meta-analyses and the range of individual scores within each study, considerable differences exist among the studies.

#### WHY DOES JUDGMENT ACCURACY DIFFER? STANDARDS, DEMANDS, AND TYPES OF JUDGMENT FORMATION

Teacher accuracy varies tremendously across and within studies. Why is that? Two moderator variables identified by Südkamp *et al.* indicate that the assumptions and methodologies used in the reported studies need to be considered, given that the variation in judgment accuracy sometimes appears to be caused by differences in study methodologies rather than in quality of teacher judgment. It obviously matters whether the domain

or aspect of student characteristics subject to teacher judgment matched with the one measured in the criterion test. According to Südkamp *et al.* “congruence in domain specificity” proved to be a moderator: Lower accuracy estimates were found in studies in which the domain specificity was incongruent (e.g., teachers rated students’ overall academic achievement, whereas students were tested on reading comprehension). Furthermore, teacher judgment accuracy was moderated by use of either informed or uninformed teacher judgments, with use of informed judgments leading to a higher correspondence between teachers’ judgments and students’ academic achievement. Judgment accuracy thus differs depending on whether or not teachers are, for example, informed about the standard of comparison or the test used for the comparisons of their judgment with student characteristics. If they are aware of the contents of the test/tasks that the students have to perform, their judgments tend to be more accurate. If, by contrast, the teachers do not know the standards or the basis for the comparison, they cannot be expected to detect information that is potentially informative about the respective characteristic or to decide whether relevant information is available. According to Funder’s (1999) realistic accuracy model, it is thus not very realistic to assume that teacher judgments are accurate (Artelt & Rausch, 2014).

Another source of the variation in judgment accuracy is the fact that different types of both teacher judgments and assessments of the corresponding student characteristics were used. The most common approach is a judgment type in which teachers are asked to judge students’ academic achievement or even ability on a rating scale (e.g., ranging from poor to excellent). Other studies use teacher judgments of individual students’ performance on particular items and compare these to the actual performance of the student on the specific tasks. Thus, judgment types differ with respect to their specificity. As will be argued throughout this essay, the quality of teacher judgments depends on which kind of judgment (e.g., global vs task-specific) teachers are asked to perform and hence on which indicators are used to estimate teacher accuracy. In what respect do these judgments differ? Let us take global versus task-specific judgments as an example to briefly illustrate differences in the demands of the judgment tasks as well as in the assumptions one has to make when comparing teacher judgments with external standards in order to assess their accuracy. Obviously, more information is available to the teachers for task-specific judgments. For global judgments, however, teachers have to infer which of the students’ specific behaviors they consider relevant for their judgment. They also have to decide which detections of specific behavior they will integrate into their overall judgment. Both tasks contain potential sources of error. For task-specific judgments, although relevant information is available in the form of information about task demands, the detection

and utilization of the available information is dependent on both the teachers' additional knowledge about the relative difficulty of the tasks (content knowledge and didactic knowledge) and their knowledge about students' individual strengths and difficulties relative to these task demands. Moreover, although the standards of comparison are quite clear for task-specific judgments, they are not necessarily so for global judgments. Given that teachers have to infer their ratings based on their definition of the respective competence domain (e.g., reading proficiency or mathematical competence), a lower accuracy of global judgments is possibly due to the lack of a common definition in this domain. While reading proficiency, for instance, is sometimes regarded mainly as a decoding skill and a kind of fluency indicator, other definitions (such as those inherent in the test used as the standard of reference in our study) are based on the ability to construct a coherent mental representation of the text. Depending on what teachers consider the key definition, their ratings may differ not because of a lack of judgment accuracy but because of a mismatch between the inherent construct of the test and the construct in the teachers' minds. For the domain of mathematics, demands and competence models are clearer. Given that mathematical skill is acquired mainly in school and that the curriculum is well defined according to age, teachers quite likely have a shared understanding of what constitutes mathematical proficiency. At the same time, this is likely to be the concept implemented in the reference tests for students, leading to higher judgment accuracy for mathematics as compared to reading (cf. Artelt & Rausch, 2014).

In this respect, teacher judgments on students' future academic success (i.e., as a basis for school transition recommendations) are global as well. They are characterized by both a certain degree of uncertainty and a lack of information about the reliability of these decisions and recommendations (for teachers but also for research). Thus, similar mechanisms related to the use of subjective standards and (social) schema in the process of judgment formation might also take place in the formation of teacher judgments on students' future educational success and attainment.

So far, accuracy of teacher judgments has been discussed with a focus on the validity of the criterion used for assessing teacher accuracy in research. In addition, the degree of specificity of teacher judgments under study has been introduced as imposing different cognitive demands on the teacher. Global judgments of students thereby seem to be most sensitive to bias (Rausch, Karing, Dörfler, & Artelt, 2015), given that the frame of reference is often unclear and the process of judgment formation relatively unstructured. Yet another source of inaccuracy (and potentially also bias) in teacher judgments is related to the fact that human judgment formation is not always as rational as it could be. Dual-process models of (social)

decision-making (Fiske, 1993) differentiate between two types of judgment formation: heuristic versus analytic.

#### HEURISTICS JUDGMENT FORMATION

Like much of human decision-making, the process of heuristic judgment formation is characterized by the use of heuristics and shortcuts, where a minimal amount of information is processed; sometimes, only a single cue is considered. Decision-making thereby relies on the activated cognitive schema or category (e.g., social categories). In the heuristic approach, student characteristics observed by the teacher are interpreted within the context of the activated category, which means that the category exerts a significant influence on the judgment process. The advantage of heuristic judgment formation is its efficiency as well as the fact that it requires little cognitive effort. Teachers use heuristics to make everyday decisions at school. However, when it comes to formal decisions, heuristic judgment formation might be a source of bias. Although little is known about the actual use of heuristics in teachers' decision-making processes, research has clearly shown that teacher judgments are indeed influenced by social categories. Relevant social categories in this regard are, for instance, gender, age, socioeconomic status, and migration status. Teachers sometimes seem to rely on common stereotypes about strengths and weaknesses in students' abilities related to these categories (Robinson-Cimpian, Lubienski, Ganley, & Copur-Gencturk, 2014). In addition, grades and school leaving certificates are influenced not only by students' objectively measured competence (e.g., via large-scale assessments of domain-specific competencies in mathematics) but also by family background characteristics (e.g., SES, Baumert, 2016).

The more strongly teachers' social categories and preconceptions affect grades and school leaving certificates, the more likely it is that students' educational pathways are shaped by bias. This applies especially to students at the middle of the achievement distribution. While teachers often find it easy to identify very high- or low-performing students, they encounter much more uncertainty when differentiating students in the middle part of the distribution. Thus, bias occurs more often for the average students, especially when teachers draw on social categories.

#### ANALYTIC TEACHER JUDGMENTS

The process of analytic judgment formation differs considerably from that of heuristic judgment formation process. Within this rational form of judgment formation, information gathering is characterized by an endeavor to take into account all necessary information (cues), which then will be considered—using optional weights for each cue—for forming the judgment



(Swets, Dawes, & Monahan, 2000). In one sense, analytic judgment formation can be understood as an optimal model of judgment formation, resulting in highly accurate judgments. Unlike the heuristic approach, however, it is a very resource-intensive process, occupying working memory capacity that cannot be used for other professional tasks. It is therefore unrealistic to assume that the teacher judgment formation process is always (or even mostly) analytic in nature.

#### WHEN AND HOW OFTEN CAN ONE EXPECT TO FIND ACCURATE JUDGMENTS?

The findings and models reported so far suggest that teachers in general are not very accurate in judging students' performance and achievements. However, in this context it is worth considering the whens and whys of decision-making. In order to gain both knowledge about the process of teacher judgment formation and a realistic picture of teachers' ability and everyday performance in making judgments, research has to take into account which ingredients are needed for accurate judgments.

#### MOTIVATION BEHIND JUDGMENTS

Dual-process models suggest that judgment formation is not always analytic and rational (see above) and that the type of judgment formation (either analytic or heuristic) depends on the motivation of the person formulating the judgment. Krolak-Schwerdt, Böhmer, and Gräsel (2013) were able to show two important findings in this regard. First, information processing of relevant cues during the phase of forming judgments about students is in fact dependent on teachers' motivation (operationalized as the relevance of the judgment to be formed). Second, teachers—but not novices—were able to change the processing mode from a heuristic to an analytic one to make relevant judgments. In the light of these and similar considerations (Kahneman, 2011), it is likely that analytic decision-making is often limited to subjectively relevant and significant diagnostic decisions. Expertise should thus not be understood as an unconditional tendency to form accurate/analytic judgments but rather as an ability to form them when needed. In addition to the motivation to consider and process all relevant cues, teachers also require a sufficient knowledge base to consider and process these cues adequately.

#### KNOWLEDGE ABOUT TASKS AND STUDENTS

Teacher judgments about student characteristics relate, more or less, to specific school subjects. The more specific the ratings are in terms of a school

subject or competence domain, the more likely it is that knowledge about the nature and demands of the respective domain is needed in order to judge students' performance effectively. Such judgments are particularly important in formative assessments that aim at improving the quality of learning. Another clear indicator of the significance of domain- or subject-specific knowledge in teacher judgments is the subject-specific cues that need to be recognized and considered if an accurate judgment is to be reached (cf. National Research Council, 2001). In order to estimate student performance accurately, teachers need to know about the epistemic demands of the respective domain: What exactly do students need to know and understand in order to be able to deliver the performance that is expected from them? On the basis of the Tetraeder model by Campione and Armbruster (1985), individual students' performance on a specific task can be understood as varying according to (i) the *characteristics of the student*, such as prior knowledge, intelligence, and motivation; (ii) *activities required* of the student, such as the application of appropriate strategies; (iii) *task demands*; and (iv) the quality and composition of *materials*.

Teachers wishing to estimate the performance of their students accurately need to take all sources into account. The more knowledge of these aspects that teachers possess, the more accurate their judgments should be. Thus, the accuracy of diagnostic judgments—either formative or summative—relies heavily on teachers' subject-specific content knowledge (task demands and knowledge about learning materials), pedagogical content knowledge (student activities needed for specific tasks), and knowledge of related student characteristics. However, surprisingly little research has assessed these knowledge components or analyzed their role within the judgment process itself—including the question of how teachers actually combine student and task variables when judging student performance.

#### KNOWLEDGE ABOUT ERRORS AND TENDENCIES

Moreover, there are good reasons to assume that teachers' knowledge about general tendencies and errors in the decision-making process—and about adequate ways of dealing with these—are a central element of a professional knowledge base and a theoretically relevant predictor of judgment accuracy. If teachers, for instance, know about the possible consequences of selective information processing (e.g., expectancy effects) and errors related to falsely relying on dominant features (e.g., halo effects) and also about contrast and anchor effects (Kahneman, 2011), they can at least reflect on their decision-making against this background.

Given the high number of formal decisions and even higher number of informal ones that teachers have to make in an everyday school context, it

is unlikely that every decision is as accurate as it could be. Teachers need to make many decisions throughout the working day “on the fly,” without the time to engage in an explicit decision-making process. Hence, not every decision can be as rational and precise as possible. At the same time, intuitive decision-making sometimes leads to valid decisions. However, little is known about when and to what extent teachers’ judgments are based on their intuition and how reliable those judgments are. Although judgments that rely on social categories and other schema are vulnerable to bias, this kind of judgment formation might be not only quick and efficient but also more reliable than expected, especially in later phases of a professional career, when schema on students (in general as well as on specific students) become more differentiated and possibly more realistic.

### DIRECTIONS FOR FUTURE RESEARCH

Teachers’ judgments are omnipresent in their professional practice, for example, as pedagogical feedback (e.g., on success or failure) or as summative judgments on students’ ability or attainment (e.g., via grades and recommendations). Students’ educational careers and psychosocial development are strongly influenced by these judgments. At the same time, there is evidence that teacher judgments are not always as valid and accurate as might be expected. Teachers obviously draw on schema and social categories during the process of judgment formation, and research on accuracy of teacher judgments (tests as criterion) shows that accuracy varies considerably within as well as between studies. It has been argued, however, that it is unrealistic to claim that teacher judgments should always be accurate, independent of the cause of judgment. It is worth considering the whens and whys of decision-making and future research should systematically take into account different judgment causes and demands. In order to gain both knowledge about the process of teacher judgment formation and a realistic picture of teachers’ ability and everyday performance in making judgments, the ingredients needed for accurate judgments need to be considered.

Up to now, surprisingly little research has assessed the role of professional knowledge needed for adequate judgments on student performance—especially related to knowledge about student variables (e.g., their dispositions and activities) as well as about task variables (e.g., demands generated by different tasks and materials). Nor has there been substantial research on the role of different knowledge facets within the judgment process itself—including the question of how teachers actually combine knowledge about students and tasks when judging student performance. Moreover, there is a research gap related to the professional process of

judgment formation. We seem to know very little about the role and use of professional knowledge in the everyday practice of judgment formation in school.

Another desideratum for future research is the need to apply adequate standards for comparisons when evaluating the quality of teacher judgments. Mixed results on the accuracy of teacher judgments can partly be attributed to the fact that research often does not have or apply adequate (valid) standards for evaluating accuracy. Given that teachers' recommendations on students' educational pathways (including school leaving certificates and grades—especially at the end of primary school) often lack feedback on students' success in secondary school, the necessity of valid benchmarks also applies to teacher judgments themselves.

Moreover, little is known about the actual use of heuristics in teachers' decision-making processes. We clearly need to know more about when and to what extent teachers' judgments are based on their intuition and how reliable those judgments are. Especially in later phases of a professional career, when schema on students (in general as well as on specific students) become more differentiated and possibly more realistic, this kind of judgment formation might be not only quick and efficient but also more reliable than expected. Given, on the one hand, findings on the detrimental consequences of biased judgments for students, especially if these are based on stereotypes, and, on the other hand, research on valid professional intuition as a result of deliberate practice (see Plessner, Betsch, & Betsch, 2010, for an overview), further elaboration is needed on the development of expertise in nonanalytic judgment formation (intuition). There is a need to better understand how teachers' professional decision-making develops as a function of educational practice and to what extent professional development is dependent on deliberate practice and on the active seeking and processing of feedback (e.g., metacognitive judgments of students regarding their own learning progress). Especially, the role of knowledge components in controlled and automatic decision-making after intensive experience on the job needs to be better understood. Attempts to reduce both social disparities and the psychosocial costs of invalid teacher judgments might benefit substantially from these findings.

## REFERENCES

- Artelt, C., & Rausch, T. (2014). Accuracy of teacher judgments. When and for what reasons? In S. Krolak-Schwerdt, S. Glock & M. Böhmer (Eds.), *Teachers' professional development: Assessment, training, and learning* (pp. 27–43). Rotterdam, The Netherlands: Sense Publishers.

- Baumert, J. (2016). Bildungsgerechtigkeit in Deutschland—ein Überblick [Educational justice in Germany—An overview]. In F. Königssteiner (Ed.), *Die Suche nach der richtigen Ordnung (Jahresband 2014)*. Frankfurt am Main, Germany: Societäts-Verlag.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften [Keyword: Professional competence of teachers]. *Zeitschrift für Erziehungswissenschaft*, 9, 469–520.
- Campione, J. C., & Armbruster, B. B. (1985). Acquiring information from texts: An analysis of four approaches. In J. W. Segal, S. F. Chipman & R. Glaser (Eds.), *Thinking and learning skills. Volume 1: Relating instruction to research* (pp. 317–359). Hillsdale, NJ: Erlbaum.
- Darley, J. M., & Fazio, R. H. (1980). Expectancy confirmation processes arising in the social interaction sequence. *American Psychologist*, 35, 867–881.
- Fiske, S. T. (1993). Social cognition and social perception. *Annual Review of Psychology*, 44, 155–194.
- Funder, D. C. (1999). *Personality judgment. A realistic approach to person perception*. San Diego, CA: Academic Press.
- Hofer, B. K. (2001). Personal epistemology research. Implications for learning and teaching. *Journal of Educational Psychology Review*, 13, 353–383.
- Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, 59, 297–313.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.
- Karing, C., Pfof, M., & Artelt, C. (2013). Is secondary school teacher judgment accuracy related to the development of students' reading literacy? In M. Pfof, C. Artelt & S. Weinert (Eds.), *The development of reading literacy from early childhood to adolescence* (pp. 279–310). Bamberg, Germany: Bamberg University Press.
- Krolak-Schwerdt, S., Böhmer, M., & Gräsel, C. (2013). The impact of accountability on teachers' assessments of student performance: A social cognitive analysis. *Social Psychology of Education*, 16, 215–239. doi:10.1007/s11218-013-9215-9.
- Looney, J. W. (2011). *Integrating formative and summative assessment: Progress toward a seamless system?* OECD education working papers, 58.
- National Research Council (2001). *Knowing what students know: The science and design of educational assessment*. Washington, D.C.: National Academy Press.
- Niegemann, H., & Stadler, S. (2001). Hat noch jemand eine Frage? Systematische Unterrichtsbeobachtung zur Häufigkeit und kognitivem Niveau von Fragen im Unterricht [Are there questions? Systematic observation of the frequency and cognitive level of questions in classes]. *Unterrichtswissenschaft*, 29, 171–192.
- OECD (2012). *Education today 2013: The OECD perspective*. Paris, France: OECD Publishing. doi:10.1787/edu\_today-2013-en.
- Plessner, H., Betsch, C., & Betsch, T. (Eds.) (2010). *Intuition in judgment and decision making*. New York, NY: Lawrence Erlbaum.
- Rausch, T., Karing, C., Dörfler, T., & Artelt, C. (2015). Personality similarity between teachers and their students influences teacher judgement of student achievement. *Educational Psychology*. doi:10.1080/01443410.2014.998629.

- Robinson-Cimpian, J. P., Lubienski, S. T., Ganley, C. M., & Copur-Gencturk, Y. (2014). Teachers' perceptions of students' mathematics proficiency may exacerbate early gender gaps in achievement. *Developmental Psychology, 50*, 1262–1281.
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz [Accuracy of teacher judgments on students' characteristics and the construct of diagnostic competence]. *Zeitschrift für Pädagogische Psychologie, 19*, 85–95.
- Stiensmeier-Pelster, J., & Heckhausen, H. (2008). Causal attribution of behavior and achievement. In J. Heckhausen & H. Heckhausen (Eds.), *Motivation and action* (pp. 349–383). New York, NY: Cambridge University Press.
- 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.
- Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1*, 1–26.
- Weinert, F. E., Schrader, F.-W., & Helmke, A. (1990). Educational expertise: Closing the gap between educational research and classroom practice. *School Psychology International, 11*, 163–180.
- Zhou, J., & Urhahne, D. (2012). Teacher judgment, student motivation, and the mediating effect of attributions. *European Journal of Psychology of Education, 28*, 1–21.

### FURTHER READING

- Artelt, C., & Rausch, T. (2014). Accuracy of teacher judgments. When and for what reasons? In S. Krolak-Schwerdt, S. Glock & M. Böhmer (Eds.), *Teachers' professional development: Assessment, training, and learning* (pp. 27–43). Rotterdam, The Netherlands: Sense Publishers.
- Plessner, H., Betsch, C., & Betsch, T. (Eds.) (2010). *Intuition in judgment and decision making*. New York, NY: Lawrence Erlbaum.

### CORDULA ARTELT SHORT BIOGRAPHY

**Cordula Artelt** is a Full Professor of Education (Educational Research) at the Otto-Friedrich University Bamberg, Germany. She holds degrees in Psychology and Education. She is a member of the Senate of the German Research Foundation (DFG) and the German Data Forum (RatSWD) and Vice Dean at the faculty of Human Sciences in Bamberg. Since 2000, she is involved in national and international organization and reporting on OECD's Programme for International Student Assessment (PISA). Her research focuses on issues related to text comprehension, metacognition, competence development, large-scale assessments, and teacher professionalism.

Website: <https://www.uni-bamberg.de/en/bildungsforschung/staff/cordula-artelt/>.

RELATED ESSAYS

- Learning Across the Life Course (*Sociology*), Jutta Allmendinger and Marcel Helbig
- Economics of Early Education (*Economics*), W. Steven Barnett
- Misinformation and How to Correct It (*Psychology*), John Cook *et al.*
- Enduring Effects of Education (*Sociology*), Matthew Curry and Jennie E. Brand
- Four Psychological Perspectives on Creativity (*Psychology*), Rodica Ioana Damian and Dean Keith Simonton
- The Organization of Schools and Classrooms (*Sociology*), David Diehl and Daniel A. McFarland
- Expertise (*Sociology*), Gil Eyal
- Cognitive Processes Involved in Stereotyping (*Psychology*), Susan T. Fiske and Cydney H. Dupree
- Controlling the Influence of Stereotypes on One's Thoughts (*Psychology*), Patrick S. Forscher and Patricia G. Devine
- Social Class and Parental Investment in Children (*Sociology*), Anne H. Gauthier
- Evolutionary Approaches to Understanding Children's Academic Achievement (*Psychology*), David C. Geary and Daniel B. Berch
- Language and Thought (*Psychology*), Susan Goldin-Meadow
- Family Relationships and Development (*Psychology*), Joan E. Grusec
- Evaluating and Rewarding Teachers (*Education*), Cassandra Hart
- An Evolutionary Perspective on Developmental Plasticity (*Psychology*), Sarah Hartman and Jay Belsky
- Multitasking (*Communications & Media*), Matthew Irwin and Zheng Wang
- Educational Testing: Measuring and Remediating Achievement Gaps (*Education*), Jaekyung Lee
- Gender Inequality in Educational Attainment (*Sociology*), Anne McDaniel and Claudia Buchmann
- Neural and Cognitive Plasticity (*Psychology*), Eduardo Mercado III
- A Bio-Social-Cultural Approach to Early Cognitive Development: Entering the Community of Minds (*Psychology*), Katherine Nelson
- Retrieval-Based Learning: Research at the Interface between Cognitive Science and Education (*Psychology*), Ludmila D. Nunes and Jeffrey D. Karpicke
- Attention and Perception (*Psychology*), Ronald A. Rensink
- Cognitive Bias Modification in Mental (*Psychology*), Meg M. Reuland *et al.*
- Curriculum as a Site of Political and Cultural Conflict (*Sociology*), Fabio Rojas

Education in an Open Informational World (*Education*), Marlene Scardamalia and Carl Bereiter

Institutional Contexts for Socioeconomic Effects on Schooling Outcomes (*Sociology*), Herman G. van de Werfhorst