American Educational Research Journal

http://aerj.aera.net

Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress

Jürgen Baumert, Mareike Kunter, Werner Blum, Martin Brunner, Thamar Voss, Alexander Jordan, Uta Klusmann, Stefan Krauss, Michael Neubrand and Yi-Miau Tsai

Am Educ Res J 2010 47: 133 originally published online 19 October 2009
DOI: 10.3102/0002831209345157

American Educational Research Journal
March 2010, Vol. 47, No. 1, pp. 133–180
DOI: 10.3102/0002831209345157

Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress

Jürgen Baumert
Mareike Kunter
Max Planck Institute for Human Development
Werner Blum
University of Kassel
Martin Brunner
University of Luxembourg
Thamar Voss
Max Planck Institute for Human Development
Alexander Jordan
University of Bielefeld
Uta Klusmann
Max Planck Institute for Human Development
Stefan Krauss
University of Kassel
Michael Neubrand
University of Oldenburg
Yi-Miau Tsai
University of Michigan

In both the United States and Europe, concerns have been raised about whether preservice and in-service training succeeds in equipping teachers with the professional knowledge they need to deliver consistently high-quality instruction. This article investigates the significance of teachers' content knowledge and pedagogical content knowledge for high-quality instruction and student progress in secondary-level mathematics. It reports findings from a 1-year study conducted in Germany with a representative sample of Grade 10 classes and their mathematics teachers. Teachers' pedagogical content knowledge was theoretically and empirically distinguishable from their content knowledge. Multilevel structural equation models revealed a substantial positive effect of pedagogical content knowledge on students' learning gains that was mediated by the provision of cognitive activation and individual learning support.

KEYWORDS: teacher knowledge, teacher education, mathematics, instruction, cognitive activation, hierarchical modeling with latent variables

Since Lee Shulman's presidential address at the 1985 American Educational Research Association meeting—in which Shulman went beyond the generic perspective of educational psychology, emphasizing the importance of domain-specific processes of learning and instruction—educational research

---

JÜRGEN BAUMERT is a co-director at Max Planck Institute for Human Development, Center for Educational Research, Lentzeallee 94, 14195 Berlin, Germany; e-mail: *sekbaumert@mpib-berlin.mpg.de*. His research interests include research in teaching and learning, cultural comparisons, large-scale assessment, and cognitive and motivational development in adolescence.

MAREIKE KUNTER is a research scientist at Max Planck Institute for Human Development, e-mail: *kunter@mpib-berlin.mpg.de*. Her research interests include teacher research, motivational processes in the classroom, and assessment of instructional processes.

WERNER BLUM is a professor of mathematics education at University of Kassel, e-mail: *blum@mathematik.uni-kassel.de*. His research interests include empirical research on instructional quality in mathematics, national and international comparison studies in mathematics, approaches to application, modeling, and proofs in mathematics instruction.

MARTIN BRUNNER is an associate professor at University of Luxembourg, e-mail: *martin.brunner@uni.lu*. His research interests include research on cognitive abilities, achievement, and achievement motivation by means of modern measurement models.

THAMAR VOSS is a predoctoral research fellow at Max Planck Institute for Human Development, e-mail: *voss@mpib-berlin.mpg.de*. Her research interests include research on instruction and learning, teacher research, and teacher beliefs.

ALEXANDER JORDAN is an academic staff member at University of Bielefeld, e-mail: *ajordan@math.uni-bielefeld.de*. His research interests include empirical research on mathematics instruction, instructional quality development, professional knowledge of mathematics teachers, and national and international comparison studies in mathematics.

UTA KLUSMANN is a research scientist at Max Planck Institute for Human Development, e-mail: *klusmann@mpib-berlin.mpg.de*. Her research interests include research in teaching and learning, teacher stress, and personal goals.

STEFAN KRAUSS is a professor of mathematics education at University of Kassel, e-mail: *skrauss@mathematik.uni-kassel.de*. His research interests include research on teaching and learning, didactics of mathematics, professional knowledge of mathematics teachers, and probabilistic reasoning.

MICHAEL NEUBRAND is a professor of mathematics instruction at University of Oldenburg, e-mail: *neubrand@mathematik.uni-oldenburg.de*. His research interests include teaching and learning mathematics in secondary school, didactics in geometry, mathematical literacy, professional knowledge of mathematics teachers, and international comparison studies in mathematics.

YI-MIAU TSAI is a research fellow at University of Michigan, e-mail: *ymtsai@isr.umich.edu*. Her research interests include research in learning and instruction, achievement motivation and goals, and context and cultural effects on motivation.

has distinguished three core dimensions of teacher knowledge: content knowledge (CK), pedagogical content knowledge (PCK), and generic pedagogical knowledge (Shulman, 1986). Various authors have added to, and further specified, these core components of teachers' professional knowledge (e.g., Grossman, 1995; Sherin, 1996; Shulman, 1987). In the research literature on teaching and teacher education, there is a shared understanding that domain-specific and general pedagogical knowledge and skills are important determinants of instructional quality that affect students' learning gains and motivational development (Bransford, Darling-Hammond, & LePage, 2005; Bransford, Derry, Berliner, & Hammerness, 2005; Grossman & McDonald, 2008; Grossman & Schoenfeld, 2005; Hiebert, Morris, Berk, & Jansen, 2007; Munby, Russell, & Martin, 2001; Reynolds, 1989). Yet few empirical studies to date have assessed the various components of teachers' knowledge directly and used them to predict instructional quality and student outcomes (Fennema et al., 1996; Harbison & Hanushek, 1992; Hill, Ball, Blunk, Goffney, & Rowan, 2007; Hill, Rowan, & Ball, 2005; Mullens, Murnane, & Willet, 1996; Rowan, Chiang, & Miller, 1997). The National Mathematics Advisory Panel (2008) summarizes the situation as follows:

> Finally, with the exception of one study that directly measured the mathematical knowledge used in teaching, no studies identified by the Panel probed the dynamic that would examine how elementary and middle school teachers' mathematical knowledge affects instructional quality, students' opportunities to learn, and gains in achievement over time. (p. 37)

In this article, we present findings from a 1-year study in which mathematical teachers' CK and PCK were assessed directly and linked to data from a comprehensive assessment of mathematics instruction and student outcomes. The core question guiding the study was whether these two components of teacher knowledge each make a unique contribution to explaining differences in the quality of instruction and student progress.

The COACTIV[1] study was conducted in Germany from 2003 to 2004 as a national extension to the 2003 cycle of the Organisation for Economic Co-operation and Development's Programme for International Student Assessment (PISA; Prenzel, Baumert, et al., 2006). The study, which had two measurement points, surveyed a nationally representative sample of Grade 10 classes and their mathematics teachers. Its objective was to investigate the implications of CK and PCK for processes of learning and instruction in secondary level mathematics (Kunter et al., 2007; Kunter, Klusmann, & Baumert, 2009). To this end, we used newly constructed knowledge tests to assess teacher knowledge directly. The teacher data were then linked to data on aspects of instruction and student outcomes. Specifically, we investigated the following hypotheses: that CK and PCK represent distinct knowledge categories, that PCK is directly associated with the quality of instruction, and that its effect on student learning is mediated by the quality of instruction.

Prior Research on CK and PCK

There is consensus in the teacher education literature that a strong knowledge of the subject taught is a core component of teacher competence (e.g., American Council on Education, 1999; Grossman & Schoenfeld, 2005; Mewborn, 2003; National Council of Teachers of Mathematics, 2000; National Mathematics Advisory Panel, 2008). Opinions on what exactly is meant by subject matter knowledge are divided, however, even for mathematics. There is disagreement on the necessary breadth and depth of teachers' mathematical training (cf. Ball & Bass, 2003; Deng, 2007; Shulman & Quinlan, 1996): Do secondary mathematics teachers need a command of the academic research knowledge taught in the mathematics departments of universities? Or is it mathematical knowledge for teaching that matters, integrating both mathematical and instructional knowledge, as taught at schools of education? There is, however, agreement among mathematics educators that "teachers must know in detail and from a more advanced perspective the mathematical content they are responsible for teaching . . . both prior to and beyond the level they are assigned to teach" (National Mathematics Advisory Panel, 2008, p. 37). What is required is a *conceptual* understanding of the material to be taught (Mewborn, 2003; National Council of Teachers of Mathematics, 2000; National Mathematics Advisory Panel, 2008).

It is thus all the more surprising that quantitative research on teacher competence is based almost exclusively on proxies such as certification status and mathematics course work completed (Cochran-Smith & Zeichner, 2005). Qualitative studies, in contrast, have closely examined the importance of a conceptual understanding of the content to be taught (Ball, Lubienski, & Mewborn, 2001; Leinhardt, 2001). In the following, we first outline the findings of quantitative studies that use distal indicators of teacher knowledge or that conceptualize CK as knowledge of high school mathematics. We then present findings from qualitative studies. Finally, we consider the findings of the research group led by Deborah Ball at the University of Michigan. This group was the first to measure elementary school teachers' mathematical knowledge for teaching directly and to examine its relationship to student progress (Hill et al., 2005; Hill et al., 2007).

Findings of Studies Using Distal Indicators

In recent years, a number of review articles have been published providing overviews of quantitative studies that have, for the most part, used distal indicators of teachers' knowledge and made no distinction between CK and PCK (Ballou & Podgursky, 2000; D'Agostino & Powers, 2009; Darling-Hammond, 2000; Floden & Meniketti, 2005; Goe, 2007; Wayne & Youngs, 2003; Wilson & Floden, 2003).

Several studies have investigated whether state *certification* as an indicator of teacher quality is reflected in enhanced student learning gains. When certification in a subject is assessed and correlated with student achievement in the same domain, findings tend to indicate a positive relationship, especially for mathematics. The most important evidence to this effect is provided by Goldhaber and Brewer's reanalyses (2000) of the National Education Longitudinal Study data and Darling-Hammond's analyses (2000) with combined data from the Schools and Staffing Survey and mathematics and reading data from the National Assessment for Educational Progress. Findings on teachers' *qualifications* (major/minor or B.A./M.A.) and *course attendance* are rather more complex. The empirical basis is provided by the work of Goldhaber and Brewer (2000), Monk (1994), and Rowan et al. (1997). Higher teacher *qualifications* tend to be associated with better student performance at secondary level, particularly in mathematics. Findings for the number of *courses* attended in the teaching subject are inconsistent across school subjects but generally positive for mathematics. Exposure to teachers who took more mathematics courses during the university-based phase of teacher training seems to have positive effects on secondary students' learning gains. Monk (1994) reported interactions with students' prior knowledge: The higher the students' prior knowledge, the more important the subject matter component of teacher training. Monk also found decreasing marginal returns on course attendance. Neither of these findings has yet been replicated. There is a clear need for studies that assess teachers' CK by means other than distal measures (National Mathematics Advisory Panel, 2008).

Findings of Studies Conceptualizing CK as Knowledge of High School Mathematics

Given the widespread agreement that teachers must know the mathematical content they are responsible for teaching not only from a more advanced perspective but beyond the level they are assigned to teach, it is surprising that several empirical studies on the impact of teacher knowledge have conceptualized teachers' mathematical CK as knowledge of high school or even elementary school mathematics. Harbison and Hanushek (1992) administered a mathematics test to fourth graders in rural areas of Brazil as well as to their teachers and used the teacher scores to predict change in students' scores in Grade 4. Mullens and colleagues (1996) used the scores attained by elementary school teachers in Belize in their final mathematics test at the end of compulsory schooling, at the age of 14 years, as an indicator of their mathematical

CK. In these studies, both indicators proved to predict student learning gains in mathematics. Drawing on data from the National Education Longitudinal Study of 1988, Rowan et al. (1997) found a positive relationship between students' learning gains and teachers' mathematical CK, as assessed by a single test item tapping teachers' knowledge of high school mathematics. Even the latest comparative international assessment conducted by the International Association for the Evaluation of Educational Achievement, the Teacher Education and Development Study in Mathematics, assesses teachers' CK at the level of advanced high school knowledge (Schmidt et al., 2007), despite attempts made in the pilot study to conceptualize it at a higher level (Blömeke, Kaiser, & Lehmann, 2008).

Findings of Qualitative Studies on the Importance of a Conceptual Understanding of Content

A considerable body of qualitative studies on the structure and effects of teacher knowledge has developed in the last 20 years, providing a rather more informative picture than that of the distal approach (Ball et al., 2001; Leinhardt, 2001; Stodolsky & Grossman, 1995). One of the major findings of qualitative studies on mathematics instruction is that the repertoire of teaching strategies and the pool of alternative mathematical representations and explanations available to teachers in the classroom are largely dependent on the breadth and depth of their conceptual understanding of the subject. Studies in which teachers were presented with examples of critical classroom events revealed that an insufficient understanding of mathematical content limits teachers' capacity to explain and represent that content to students in a sense-making way, a deficit that cannot be offset by pedagogical skills. Ball (1990) and Ma (1999) demonstrated this relationship for multiplication and place values; Borko et al. (1992) and Simon (1993), for division; Even (1993), Stein, Baxter, and Leinhardt (1990), and Heaton (2000), for patterns and functions; and Putnam, Heaton, Prawat, and Remillard (1992), for geometry. Given their case studies, Putnam et al. concluded that the efforts of teachers with a limited conceptual understanding "fell short of providing students with powerful mathematical experiences" (p. 221).

In her comparison of teachers in China and the United States, Ma (1999) showed that a "profound understanding of fundamental mathematics" is reflected in a broad repertoire of pedagogical strategies over a range of mathematical topics. The breadth, depth, and flexibility of Chinese teachers' understanding of the mathematics they teach afford them a broader and more varied repertoire of strategies for representing and explaining mathematical content than what is available to their colleagues in the United States. Intervention studies show that enhancement of mathematical CK *can* lead to higher-quality instruction (e.g., Fennema & Franke 1992; Swafford, Jones, &

Thornton, 1997). Qualitative research on teacher knowledge acknowledges that the mathematical CK required for high-quality instruction is not general mathematical knowledge that is picked up incidentally but profession-specific knowledge that is acquired in university-level training and can be cultivated through systematic reflection on classroom experience (Ball et al., 2001; Berliner, 1994; Grossman, 2008).

Qualitative Studies Distinguishing Between CK and PCK

CK is not a panacea, however. Findings show that CK remains inert in the classroom unless accompanied by a rich repertoire of mathematical knowledge and skills relating directly to the curriculum, instruction, and student learning. The case study by Eisenhart et al. (1993) brought fame to Ms. Daniels, who had a reasonable conceptual understanding of the division of fractions but was unable to present her students with a correct mathematical representation of the problem. Similar findings have been reported for rates (A. G. Thompson & P. W. Thompson, 1996; P. W. Thompson & A. G. Thompson, 1994) and for multiplication (Ball, 1991). These findings are complemented by case studies of specific instructional episodes, which show that teachers with equivalent levels of subject matter knowledge may differ considerably in their pedagogical repertoire and skills depending on their teaching experience (Schoenfeld, 1998; Schoenfeld, Minstrell, & van Zee, 2000). PCK thus seems to vary—at least to a certain degree—independently of CK and to be a knowledge component in its own right.

In the words of Kahan, Cooper, and Bethea (2003), strong mathematical CK seems to be "a factor in recognizing and seizing teachable moments" (p. 245), but it does not guarantee powerful mathematical experiences for students. What is required here is PCK, "which involves bundles of understandings that combine knowledge of mathematics, of students, and of pedagogy" (Ball et al., 2001, p. 453). According to Ball and colleagues (2001), it is PCK in particular that underlies the development and selection of tasks, the choice of representations and explanations, the facilitation of productive classroom discourse, the interpretation of student responses, the checking of student understanding, and the swift and correct analysis of student errors and difficulties. In summary, findings suggest that—in mathematics at least—a profound understanding of the subject matter taught is a necessary, but far from sufficient, precondition for providing insightful instruction (see also Borko & Livingston, 1989; Kahan et al., 2003).

Mathematical Knowledge for Teaching

Drawing on both CK and PCK, the research group headed by Deborah Ball at the University of Michigan has developed a theoretical framework and set of measurement instruments for the assessment of elementary school teachers' mathematical knowledge for teaching (Ball & Bass, 2003; Hill, 2007; Hill, Schilling, & Ball, 2004). Ball and colleagues see mathematics teachers' professional CK as the mathematics they need to know in order to teach effectively. On this basis, they distinguish common knowledge of content (the mathematical everyday knowledge that all educated adults should have) from specialized knowledge of content (the specialist knowledge acquired through professional training and classroom experience; Hill et al., 2004; Schilling & Hill, 2007). They further distinguish a third dimension of mathematical knowledge, which links mathematical content and student thinking (typical errors or student strategies)—namely, knowledge of students and content. Three content areas of elementary school mathematics are distinguished: numbers/operations, patterns/functions, and algebra. The Michigan group used a matrix of these content areas and knowledge dimensions as a theoretical structure for developing test items, in which items were allocated to individual cells of the matrix on the basis of a priori theoretical considerations.

We can gain insights into the Michigan group's conceptualization of its knowledge dimensions by taking a closer look at its descriptions of these dimensions and published test items. First, items developed to tap common knowledge of content draw on everyday knowledge ("What is the number halfway between 1.1 and 1.11?"; "Can the number 8 be written as 008?") as well as on typical secondary school knowledge that is often lost after school ("What power of ten equals one?"; Hill et al., 2004). Second, the authors see teachers' specialized knowledge of content as both a conceptual understanding of topics typically taught at school ("Show that any number is divisible by 4 if the number formed by its last two digits is divisible by 4") and the specific knowledge required for teaching, "including building or examining alternative representations, providing explanations, and evaluating unconventional student methods" (Hill et al., 2004, p. 16)—that is, knowledge that could also be classified as PCK (Ball et al., 2001). Third, items tapping knowledge of students and content assess diagnostic skills, such as evaluating the diagnostic potential of tasks or recognizing typical student errors (Hill et al., 2004; Schilling & Hill, 2007)—another facet of PCK.

Hill et al. (2004) and Schilling (2007) tested their theoretical model with a large sample of teachers at California's Mathematics Professional Development Institutes. The empirical findings did not provide support for the complex structure of the model. Rather, exploratory factor analyses suggested a model with three factors: two content factors (numbers/operations and patterns/functions/algebra) and a student factor that seems to tap diagnostic skills. Common knowledge and specialized knowledge of content were not empirically distinguishable. Given these analyses, the authors decided to develop a unidimensional item response theory–scaled test to assess elementary school teachers' mathematical knowledge for teaching that included both common knowledge items and specialized knowledge items but no items tapping students' thinking. This test thus assesses an amalgam of the mathematical everyday knowledge needed by adults, the conceptual understanding of mathematical topics typically taught at elementary school, and the mathematical knowledge relating directly to the instructional process (PCK), excluding diagnostic skills.

Hill et al. (2005) examined the predictive validity of their competence measure by using the item response theory score to predict elementary students' learning gains. They drew on a sample of schools participating in three Comprehensive School Reform programs and a matched group of control schools. Multilevel analyses showed that elementary teachers' mathematical knowledge for teaching indeed predicted students' learning gains in two different grades; in fact, the effect was practically linear. This study provided the first conclusive evidence for the practical importance of teachers' mathematical knowledge in terms of both the mathematical knowledge that adults use in everyday life and the specialized knowledge that teachers use in classrooms. From a video study of 10 teachers, Hill and colleagues (2007) presented more qualitative data indicating that mathematical knowledge for teaching as assessed by the Michigan group is also associated with the mathematical quality of instruction. Thus, this test seems to provide a good overall assessment of mathematical knowledge for teaching, but it does not test the implications of CK and PCK for instruction. Insights into these mechanisms are particularly important for the design of teacher-training programs, however.

The Present Study: Theoretical Framework and Research Questions

What kind of subject matter knowledge do teachers need to be well prepared for their instructional tasks? To what degree does their mastery of the content influence their instructional repertoire? To address these questions, we need to draw a theoretical and empirical distinction between CK and PCK and examine their implications for teaching and learning. The present study therefore investigates specific samples of teachers who—as a result of their different training—can be expected to differ substantially in terms of both CK and PCK.

Conceptualizing and Assessing CK and PCK Separately

The present research evolved as part of the COACTIV study conducted at the Max Planck Institute for Human Development, Berlin, in collaboration with mathematicians at the universities of Kassel, Bielefeld, and Oldenburg. COACTIV investigates the professional knowledge of secondary school mathematics teachers. The objective is to conceptualize and measure CK and PCK separately, to determine the implications of each component of teachers' professional knowledge for processes of instruction and learning (Krauss, Brunner, et al., 2008; Kunter et al., 2007).

The COACTIV framework draws on the work of Deng (2007), Goodson, Anstead, and Mangan (1998), Shulman and Quinlan (1996), and Stengel (1997), who have pinpointed the specific and, at the same time, related nature of knowledge embedded in academic disciplines and school subjects. This approach clearly differs from conceptions that localize teacher CK at the level of advanced high school knowledge (e.g., Schmidt et al., 2007). For reasons of theoretical clarification, Krauss, Baumert, and Blum (2008) proposed the following hierarchical classification of mathematical knowledge: (a) the academic research knowledge generated at institutes of higher education, (b) a profound mathematical understanding of the mathematics taught at school, (c) a command of the school mathematics covered at the level taught, and (d) the mathematical everyday knowledge that adults retain after leaving school. We conceptualize mathematics teachers' CK as the second type: a profound mathematical understanding of the curricular content to be taught. This approach is in line with the National Council of Teachers of Mathematics (2000) and the National Mathematics Advisory Panel (2008), which refer to a conceptual understanding of the mathematics to be taught. CK required for teaching has its foundations in the academic reference discipline, but it is a domain of knowledge in its own right that is defined by the curriculum and continuously developed on the basis of feedback from instructional practice (Deng, 2007; Goodson & Marsh, 1996; Mitchell & Barth, 1999).

Moreover, the COACTIV framework assumes that this CK is theoretically distinguishable from PCK, which forms a distinct body of instruction- and student-related mathematical knowledge and skills—the knowledge that makes mathematics accessible to students. From suggestions by Shulman (1986), Krauss et al. (2008) distinguished three dimensions of PCK: knowledge of mathematical tasks as instructional tools, knowledge of students' thinking and assessment of understanding, and knowledge of multiple representations and explanations of mathematical problems (cf. Ball et al., 2001). These three components were derived by considering the demands of mathematics instruction:

1. One defining characteristic of mathematics instruction is that it is choreographed through the teacher's selection and implementation of tasks and activities. Tasks and subsequent task activities create learning opportunities and determine the internal logic of instruction, the level of challenge, and the level of understanding that can be attained (De Corte, Greer, & Verschaffel, 1996; Hiebert et al., 2005). Knowledge of the potential of mathematical tasks to facilitate learning is thus a key dimension of PCK.
2. Teachers have to work with students' existing beliefs and prior knowledge. Knowledge of student beliefs (misconceptions, typical errors, frequently used

strategies) and the ability to diagnose students' abilities, prior knowledge, knowledge gaps, and strategies are thus a core component of PCK. Errors and mistakes in particular provide valuable insights into students' implicit knowledge (Vosniadou & Vamvakoussi, 2005; Vosniadou & Verschaffel, 2004).

3. Knowledge acquisition and, especially, the achievement of a deep understanding of mathematical content are active processes of construction. These processes require guidance and support, however (Mayer, 2004; Sfard, 2003), particularly when comprehension problems occur. One of the ways in which teachers can support students' mathematical understanding is by offering multiple representations and explanations.

Given these theoretical considerations, the COACTIV group has developed a CK test to assess teachers' deep understanding of the mathematical content covered in secondary school, as well as a separate PCK test to assess their knowledge of tasks, student ideas, and representations and explanations. As Krauss, Brunner, et al. (2008) have shown, confirmatory factor analyses support the theoretically postulated two-factor structure of mathematics teachers' subject matter knowledge. The two factors show substantial intercorrelation (r_{latent} = .79) that increases as a function of teachers' expertise but are clearly distinguishable. For each test, a one-dimensional two-parameter item response theory model provides a good fit to the empirical data. The technical details of the tests are presented in the Method section. In a validation study with samples of advanced university students majoring in mathematics, certified mathematics teachers, biology and chemistry teachers, and high school students in advanced placement courses in mathematics, Krauss, Baumert, and Blum (2008) found empirical evidence for the classification of mathematical CK as proposed above as well as for the conception of CK and PCK as *specific* facets of teachers' professional knowledge.

Systematic Variation in CK and PCK: An Effect of Teacher Training?

As forms of domain-specific professional knowledge, CK and PCK are thought to be acquired through formal training at the university level, supervised internships, and reflected teaching experience and not picked up incidentally. We expect the level of knowledge attained to depend on the length, intensity, and quality of the teacher-training program attended. Germany offers a unique natural experiment for testing the relationship between teachers' CK/PCK and type of training. Whereas in the United States there is much variety in preservice training (Grossman & McDonald, 2008; Zeichner & Conklin, 2005), thereby making it difficult for researchers to define treatments and provide evidence for effects, teacher education in Germany is standardized by the state. Moreover, universities and colleges offer two distinct teacher education programs, corresponding with the tracking system implemented after Grade 4. In almost all German states, students approaching the end of Grade 4 are assigned to different secondary tracks—usually in separate schools—on the basis of their performance to date. The number of tracks implemented ranges from two (Gymnasium, Sekundarschule) to four (Gymnasium, Realschule, Hauptschule, Gesamtschule). In all states, however, a clear distinction is made between the academic track (Gymnasium) and the nonacademic tracks. This distinction is reflected in the structure of teacher training. Teacher candidates aspiring to teach at secondary level must choose between degree programs, qualifying them to teach in either the academic track (Gymnasium; 5-year training program) or the other secondary tracks (e.g., Realschule or Sekundarschule; 4-year training program). The university curricula are state regulated, with the basic structure being determined at federal level, and the two program types differ in certain ways across all states. In terms of subject matter courses, the requirements for the academic track (Certification Type 1) are on average one third higher than those for the nonacademic tracks (Certification Type 2). In terms of teaching methods courses, requirements are similar for both certification types. If at all, they are higher in nonacademic track programs. Teacher candidates for the academic track attend the same subject matter courses as students majoring in mathematics, in mathematics departments. Teacher candidates for the nonacademic tracks attend subject matter courses taught by mathematics educators in mathematics departments or in schools of education, who are also responsible for the methods courses in *both* types of program. Having completed the university-based part of their training, all teacher candidates enter a structured and supervised induction program lasting 18–24 months.

A third group of teachers was certified in the former German Democratic Republic (Certification Type 3), having attended training programs lasting 4 years (until 1982) or 5 years (1983–1989). In many cases, courses covered both subject matter and teaching methods, making it impossible to quantify the amount of time allocated to each. These integrated training programs included an internship of several months.

If CK, as we conceptualize it, is indeed dependent on the type of training program attended, teachers of the different certification types can be expected to differ markedly in their CK, even when selective intake to the programs is controlled. Specifically, we expect Type 1 teachers to outscore their colleagues of the other types. It is harder to make predictions for PCK. Because we hypothesize CK to be a necessary condition for PCK, we expect Type 1 teachers to outscore Type 2 teachers in PCK, but the differences between the two groups to be much smaller than for CK. We do not expect these differences in PCK in favor of Type 1 teachers to persist when CK is controlled. Given the integrated nature of their training, no predictions can be made with respect to the PCK scores of teachers certified in the former German Democratic Republic (Type 3).

The Differential Implications of CK and PCK for Teaching and Student Progress

PCK is inconceivable without CK. Given the findings of the qualitative case studies described above, however, we assume that PCK is needed over and above CK to stimulate insightful learning. We address our hypothesis that CK and PCK have differential implications for teaching and learning by testing the predictive significance of the two knowledge constructs (Hill et al., 2007).

As Seidel and Shavelson (2007) have pointed out, the reviews and meta-analyses summarizing the state of research on learning and instruction over the last two decades differ considerably in terms of the labeling and categorization of teaching variables. It is thus difficult to compare effect sizes. Nevertheless, although the terminology differs, three components of instruction have emerged consistently as being crucial for initiating and sustaining insightful learning processes in mathematics lessons (Brophy, 2000; Helmke, 2009; Scheerens & Bosker, 1997; Seidel & Shavelson, 2007; Shuell, 1996; Walberg & Paik, 2000; Walshaw & Anthony, 2008). These three components are as follows: cognitively challenging and well-structured learning opportunities; learning support through monitoring of the learning process, individual feedback, and adaptive instruction; and efficient classroom and time management.

In mathematics lessons, the level of cognitive challenge is determined primarily by the type of problems selected and the way they are implemented. Cognitively activating tasks in the mathematics classroom might, for example, draw on students' prior knowledge by challenging their beliefs. Cognitive activation may also be prompted by class discussion if a teacher does not simply declare students' answers to be "right" or "wrong" but encourages students to evaluate the validity of their solutions for themselves or to try out multiple solution paths. It is often the implementation of tasks in the classroom that trivializes cognitively challenging problems, turning them into routine tasks (Stigler & Hiebert, 2004). Another facet of cognitively activating instruction is the fit between the topics and materials chosen by the teacher and the curricular demands of the grade or course. Instructional alignment ensures that the instruction provided corresponds with the level specified in the curriculum (Attewell & Domina, 2008). This is particularly important in continental European countries, where curricula with compulsory subject content are mandated by the state.

The second dimension of high-quality instruction considered in COACTIV is the individual learning support provided by the teacher. Studies based on motivational theories show that simply providing students with challenging tasks is not enough to motivate them to engage in insightful learning processes; rather, they need to be supported and scaffolded in their learning activities (Pintrich, Marx, & Boyle, 1993; Stefanou, Perencevich, DiCintio, & Turner, 2004; Turner et al., 1998). The ongoing monitoring of difficulties and calibrated support that addresses students' difficulties while respecting their autonomy not only foster students' motivation but are essential components of powerful learning environments in terms of cognitive outcomes (Greeno, Collins, & Resnick, 1996; Puntambekar & Hübscher, 2005).

The third crucial dimension of instruction is that of classroom management. Given the complex social situation of the classroom, where interpersonal conflicts and disruptions are an everyday reality, it is crucial to ensure sufficient learning time by establishing and maintaining structure and order. Efficient classroom management—that is, preventing disruption and using classroom time effectively—is a robust predictor of the quality of instruction and of students' learning gains (Seidel & Shavelson, 2007, p. 481; Walberg & Paik, 2000; Wang, Haertel, & Walberg, 1993).

In our longitudinal extension of PISA 2003 (COACTIV), we drew on student and teacher ratings, as well as a sample of instructional materials, to assess these three components of instruction (Kunter et al., 2007). Kunter et al. (2006) and Dubberke, Kunter, McElvany, Brunner, and Baumert (2008) showed that the three-component model provided a good fit to the empirical data and substantially predicted student progress.

In the present study, we assume teachers' professional knowledge to be an important resource in facilitating the provision of varied, challenging, and motivating learning opportunities. Specifically, we expect PCK to be a major prerequisite for instruction that is both cognitively activating and adaptive, with teachers responding constructively to student errors and providing individual learning support; furthermore, we expect CK to be necessary for, but not identical with, a rich repertoire of skills and methods for teaching mathematics. This higher instructional quality should be reflected in higher student learning progress. We thus propose a mediation model, in which the positive effect of PCK on students' learning gains is mediated by the provision of cognitively challenging learning opportunities and individual learning support. If the hypothesis formulated on the basis of the qualitative studies on teacher knowledge holds and CK alone is not a sufficient precondition for the provision of powerful learning environments, the mediation model estimated for PCK should not apply to CK, or it should apply to only a limited extent: The relations between CK and cognitive activation/individual learning support are expected to be statistically significantly lower than those estimated for PCK.

To test the discriminant validity of our assessments of CK and PCK, we include classroom management in our model, which should vary independently of mathematical knowledge.

Moderation Hypothesis: Interaction of PCK and Secondary Track

Given a reanalysis of the mathematics data from the Tennessee Class Size Experiment (Project STAR; Nye, Hedges, & Konstantopoulos, 2000), Nye, Konstantopoulos, and Hedges (2004) showed that in the first 3 years of elementary schooling, the variance in student learning gains that is attributable to teacher effectiveness is larger in low socioeconomic status (SES) schools than in high SES schools. In other words, the teacher assigned to a class matters more in low SES schools than in high SES schools. The authors interpret this finding as reflecting a creaming effect: High SES schools are able to recruit better teachers and thus develop a more homogeneous staff (Zumwalt & Craig, 2005). We expect this moderator effect to occur systematically in tracked systems, such as those implemented in the German-speaking countries.

If teachers' CK and PCK scores depend on the type of training program attended, then the CK and PCK of mathematics teachers in the lower nonacademic tracks can be expected to be lower than that of their colleagues in the academic track. Moreover, if the quality of mathematics instruction is indeed dependent on PCK, an interaction effect of PCK and secondary track can be expected for student progress. In this case, teacher knowledge should be particularly important for the learning gains of weaker students.

Method

Study Design

COACTIV was conceptually and technically embedded in the German extension to the 2003 cycle of the Organisation for Economic Co-operation and Development's PISA study (2004), which extended the international cross-sectional design involving an age-based sample of 15-year-olds to a grade-based study spanning a 1-year period from the end of Grade 9 to the end of Grade 10. Grade 10 students in Germany are the same age as their counterparts in the United States. In some German states, however, students in the lowest secondary track (Hauptschule) graduate and begin occupational training after Grade 9. As such, our study population covers about 80% of the age cohort and is thus more selective than what it would be for cohorts of freshmen in U.S. high schools.

Only classes in the school types that implement a 10th grade nationwide participated in the longitudinal extension. At the end of Grades 9 and 10, students in the PISA classes were administered achievement tests, as well as questionnaires assessing background data and aspects of their mathematics instruction. Within the framework of COACTIV, the mathematics teachers of these PISA classes were administered tests on facets of their professional knowledge, as well as questionnaires regarding their social and occupational background, motivational aspects of the teaching profession, beliefs on teaching and learning, perceptions of their own instruction, and professional self-regulation skills (Kunter et al., 2007).

Sample

A sample of Grade 9 students and their mathematics teachers that was representative for Germany was drawn in a two-step sampling process. First, a random sample of schools stratified by state and school type was drawn proportional to size. Second, two intact Grade 9 classes and their mathematics teachers were selected at random in each school. The coverage rate was 100% at the school level, 95% at the student level, and 94% and 84% at the teacher level in the first and second wave of assessment, respectively (Carstensen, Knoll, Rost, & Prenzel, 2004). A total of 181 teachers with 194 classes and 4,353 students participated in the longitudinal study (13 of the participating teachers taught parallel classes in the same school). All teachers had studied mathematics at university level and were licensed to teach the subject. Teacher candidates in Germany are required to study two teaching subjects at college. Of the teachers in our sample, 147 had majored in mathematics and 34 had minored in the subject; 26% had completed a 5-year training program for the academic track (Certification Type 1); 43%, a 4-year program for the nonacademic track (Certification Type 2); and 31%, a 4-year integrated training program in the former German Democratic Republic (Certification Type 3).[2] Unsurprisingly, 84% of teachers with Certification Type 1 taught in the academic track; 96% of teachers with Certification Type 2 taught in the nonacademic track. Teachers of Certification Type 3 taught in both tracks (66% nonacademic, 34% academic). With a mean age of 48 years, most of the teachers in the sample (48% women) looked back on many years of classroom experience (M = 22 years). Mean class size was 24 students. Analysis of sampling bias at the student level showed that the longitudinal sample can be considered representative of the 10th grade, which excludes Hauptschule students (Prenzel, Carstensen, Schöps, & Maurischat, 2006). Likewise, findings can be generalized to the population of mathematics teachers teaching in Grade 10 classrooms in Germany.

Measures

Teachers. The paper-and-pencil test used to assess CK consisted of 13 items covering arithmetic (including measurement; 4 items), algebra (2 items), geometry (1 item), functions (1 item), probability (1 item), and geometry, functions, and algebra (4 items). An additional item tapping probability was excluded due to poor model fit. All items required complex mathematical argumentation or proofs. The items covered mathematical topics that are compulsory for Grades 5 to 10 and that are particularly appropriate for assessing the conceptual understanding of mathematical content. A sample item is presented in Figure A1 of the appendix. The reliability of the total test was $r_{\mathrm{KR20}} = .83$ (Krauss, Brunner, et al., 2008).

Three facets of mathematics teachers' PCK were assessed. The *tasks* dimension assessed teachers' ability to identify multiple solution paths (4 items). The *students* dimension assessed their ability to recognize students' misconceptions, difficulties, and solution strategies. To this end, teachers were presented with classroom situations and asked to detect, analyze, or predict typical student errors or comprehension difficulties (7 items). Finally, the *instruction* dimension assessed teachers' knowledge of different representations and explanations of standard mathematics problems. Teachers were presented with 10 vignettes of typical classroom situations and asked to suggest as many ways as possible of supporting insightful learning per situation (10 items). Most items were partial-credit items. The reliability of the total test was $r_{\mathrm{KR20}} = .78$ (Krauss, Brunner, et al., 2008). Figure A1 also presents a sample item for each facet of PCK.

For each test, approximately twice as many items were first trialed in individual interviews and then piloted in a separate sample. The tests were then optimized in terms of psychometric quality, content validity, and test time. The content validity of the two final tests was rated by experts. To ensure that the tests really assess knowledge specific to mathematics teachers, we administered both tests to small samples of (a) high school students in advanced mathematics courses and (b) science teachers who had not studied mathematics. Both groups found the items of both tests practically unsolvable (Krauss, Baumert, et al., 2008).

For reasons of validity, all questions were open-ended; no multiple-choice items were used (Schoenfeld, 2007). The tests were conducted by trained test administrators in standardized single-interview situations as power tests without time limits. The mean total time needed to complete both tests was about 2 hours, which proved to be the maximum reasonable test time. The use of calculators was not permitted, because it would have compromised the validity of some items.

All items were coded independently by two trained raters using a standardized manual. In the case of rater disagreement, consensus was reached by means of discussion. Interrater agreement was satisfactory, with a mean $\rho = 0.81$ and $SD = 0.17$ (Brennan, 2001; Shavelson & Webb, 1991). The tests were item response theory–scaled using PARSCALE 4.1 (Scientific Software International, Lincolnwood, IL); a two-parametric partial-credit model was applied. The model fit for both tests was good. The person parameters are weighted likelihood estimates.

Instruction. The three aspects of instruction were assessed via different data sources. We assessed the provision of cognitively activating learning opportunities using a newly developed domain-specific approach that reconstructs learning situations at the task level. Specifically, participating teachers were asked to submit all tests and examinations they had set in the school year, as well as samples of homework assignments and tasks used to introduce two compulsory topics in Grade 10 mathematics. All tasks were compiled in a database and categorized by trained mathematics students using a classification scheme specially developed for COACTIV (Jordan et al., 2006). Pilot studies showed that the tasks set in tests and examinations, in particular, provide a valid reflection of the task structure found in instruction. In Germany, teachers are obliged to set tests four to six times per school year at the end of each instructional unit. These tests are always developed by the teachers themselves. Practically all questions have an open-ended format; multiple-choice questions are rare. The content, structure, and demands of these tasks reflect the teachers' expectations of their students. In accordance with the classification system developed by Jordan et al. (2006), the cognitive demands of the test and examination tasks were coded on three dimensions: type of mathematical task (three levels: purely technical, computational modeling, conceptual modeling), level of mathematical argumentation required (four levels: no argumentation required, low, intermediate, high level of argumentation), and translation processes within mathematics (four levels: no translation required, low, intermediate, high level of translation). The mean score across all test tasks submitted for the school year was used in the further analyses. On average, each teacher submitted 53 tasks.

The curricular level of tasks was used as an additional indicator of cognitive activation. To this end, all test and examination tasks were coded by curriculum experts in terms of their correspondence with the Grade 10 curriculum (low alignment: knowledge of elementary mathematics; moderate alignment: knowledge of simple junior high school mathematics; high alignment: knowledge of advanced junior high school mathematics). The mean score across all tasks was used in the further analyses. Figure A2 provides examples of tasks of differing cognitive demands and curricular levels.

The second dimension of instructional quality, individual learning support, was operationalized by six student rating scales, each comprising three to four items. The scales tapped the degree to which teachers provided adaptive explanations, responded constructively and patiently to errors, whether students perceived the pacing as adequate, and whether the teacher–student interaction was respectful and caring (see Table A1 for item examples). The intraclass correlation coefficient (ICC) taking into account the number of student raters per class (ICC_2) was used as a reliability measure (Lüdtke, Trautwein, Kunter, & Baumert, 2006). Overall, ICC_2 above .82 indicated good reliability of the student responses aggregated at class level (see Table A1 for reliabilities and further notes on ICC_1 and ICC_2).

Effective classroom management was operationalized by scales tapping student and teacher perceptions. Agreement between teacher and student judgments was high (Kunter & Baumert, 2006). General disciplinary climate was measured by an eight-item scale tapping teachers' perceptions of their classrooms (Cronbach's $\alpha = .82$). The two other indicators, which addressed the prevention of disruption and effective use of time, were each operationalized by a three-item scale tapping student perceptions. With an ICC_2 of .89

and .90, respectively, the reliability of the class-mean student ratings was very high (see Table A1 for sample items and reliabilities).

Students. Mathematics achievement was assessed at the end of Grade 10 by a test covering the standard content stipulated in the federal states' curricula for Grade 10 mathematics. Rasch scaling was conducted using ConQuest 2.0 (Assessment Systems, St. Paul, MN), and the partial credit model was used for all analyses (Wu, Adams, & Wilson, 1997). The reliability of the full test was $r_{\mathrm{KR20}} = .79$. To derive latent estimates of mathematics achievement in multilevel structural equation models, we split the test into two parts at random. Two achievement scores per person are thus available as weighted likelihood estimates (Warm, 1989).

The PISA literacy tests (Organisation for Economic Co-operation and Development, 2004) were used to assess mathematics and reading literacy at the end of Grade 9. Both tests, which are Rasch scaled, use a multimatrix design to ensure a broad coverage of content domains (Carstensen et al., 2004). Their reliability was $r_{\mathrm{KR20}} = .93$ and $r_{\mathrm{KR20}} = .88$, respectively. Weighted likelihood estimates were used for the person parameters. Mental ability was assessed by two subtests of the Cognitive Ability Test that tap verbal and figural reasoning and are regarded as markers of fluid intelligence (Heller & Perleth, 2000). All students were tested at the end of Grade 9. The subtests were Rasch scaled together using ConQuest; weighted likelihood estimates were again used for the person parameters (Wu et al., 1997). Reliability was $r_{\mathrm{KR20}} = .88$. The social status of the students' families was operationalized by the International Socio-Economic Index, which was developed by Ganzeboom and Treiman (2003) on the basis of the International Labour Office's occupation classification system. Both parents' most recent occupations were compared, and that with the higher status was used in the analyses. The family's educational background was measured by six hierarchically ordered levels of qualification that were dummy coded (Baumert & Schümer, 2001). Migration status was defined in terms of the parents' country of birth. If at least one parent was born outside Germany, the family was classified as immigrant. Parental occupation, education, and immigration status were assessed by a parent questionnaire.

Data Analyses

Statistical model. This study uses a mediation model to test the extent to which CK and PCK influence instructional quality, in turn affecting students' learning gains in mathematics. To this end, the study capitalizes on the naturally occurring variation in instructional quality between classes. The allocation of students to classes, and of classes to teachers, does not occur at random, however. As outlined above, not only are students allocated to tracks on the basis of their aptitude and achievement, but teachers in the academic-versus-nonacademic tracks differ in their training and certification. As a result, the comparison groups are not equivalent, and teacher characteristics covary with school type. For treatment effects to be properly estimated, it is thus vital that the nonobserved assignment process be correctly specified (Rosenbaum & Rubin, 1983; Winship & Morgan, 2007).

For students, the transition from elementary schooling to the tracks of the secondary system is highly institutionalized. Students are allocated to secondary tracks on the basis of a teacher recommendation, which is based on their German and mathematics achievement as well as an assessment of their general ability and aptitude. The teacher's recommendation tends to be binding. Empirical studies show that parental decisions deviating from this recommendation can be explained by the family's SES and education. Given data on a student's achievement in German and mathematics, mental ability, and social and ethnic background characteristics, it is possible to predict with a high degree of certainty the secondary school track attended (Ditton, 2007). Parents' freedom to choose between schools of the same track is limited in Germany by the differentiated structure of the secondary system (all tracks must be accessible within a reasonable distance) and by the low proportion of private schools. As a rule, parents choose the school nearest their home. To account for this assignment process, we used the following variables obtained at the Grade 9 assessment to control for selective intake to school types and classes at the individual level: prior knowledge of mathematics, reading literacy, mental ability, parental education, social status, and immigration status.[3]

Teachers' allocation to schools of the academic or nonacademic track is determined by their university training. Within tracks, teachers are centrally allocated to schools on the state level; the schools have no formal say in the decision. The decisive criterion is the fit between the teaching subjects required at the school and the teachers' combination of teaching subjects (major/minor). If there is more than one teacher with a suitable profile in the central applicant pool, the overall grade awarded at state certification is decisive. This procedure results in a quasi-randomized allocation of teachers to schools within tracks. To control for selective access to the different teacher education programs, we asked teachers to report their final high school grade point average on a categorical scale from 1 (highest) to 6 (lowest).

We used multilevel structural equation models with latent variables for our analyses. We specified a two-level model, controlling for selective intake to classes at the individual level and then investigating the influence of teachers' professional knowledge on instructional quality and students' learning outcomes at the class level. The dependent variable—student achievement at the end of Grade 10—was modeled as a latent construct indicated by the scores on the two parts of the mathematics achievement test as described above. On the class level, CK, PCK, and the dimensions of instructional quality were specified as latent constructs at the class level. For CK, the 13 test items were divided into two parcels, which served as indicators of the latent factor. For PCK, the three subscales were the manifest indicators for the latent construct. The dimensions of instructional quality assessed by student reports were conceived as hierarchical factors and modeled on the individual and class level simultaneously. The covariates at individual level were manifest variables. In addition, we controlled for the academic track of the class.

We specified the model presented in Figure 1, which was estimated separately for PCK and CK. We then compared the structural parameters of the

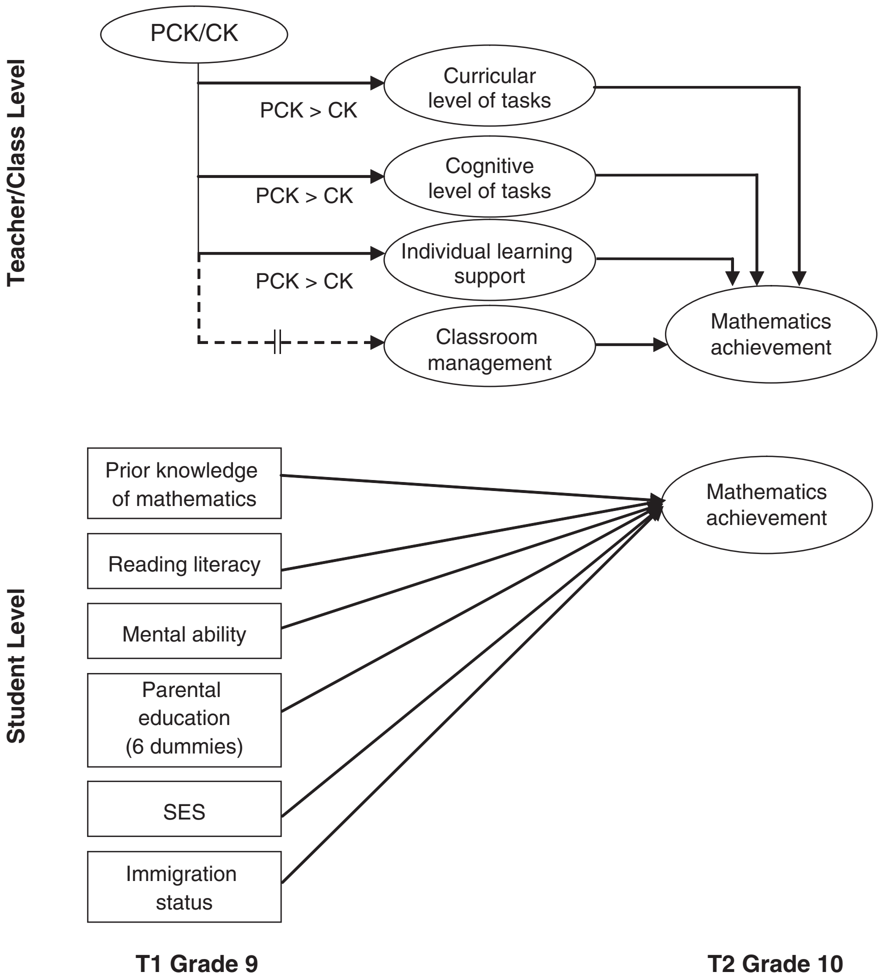


Figure 1. **Hierarchical linear model to test the significance of teachers' pedagogical content knowledge (PCK) and content knowledge (CK) for instructional quality and student learning.**

Note. SES = socioeconomic status; - - - = no relation expected.

two models at the class level in terms of their correspondence with the predictions of the theoretical framework.

All analyses were conducted with M*plus* 4.01 (Muthén & Muthén, 2004). We report several goodness-of-fit measures: chi-squre, comparative fit index, root mean square error of approximation, and standardized root mean square residual (Bollen & Long, 1993). Effect sizes were computed using the model constraints module implemented in M*plus*.

Missing values. Missing values are a widespread problem in longitudinal studies. Cases with missing values are often removed from data sets (listwise or pairwise deletion) or replaced by mean values. Both these approaches assume that data are "missing completely at random." There is now consensus that multiple imputation or the full information maximum likelihood estimator, both of which assume only that data are "missing at random," are preferable methods for dealing with missing values (Peugh & Enders, 2004). Although the amount of missing data in our study was relatively small, its management merits careful consideration. The percentage of missing values differed across assessment domains. Mathematics scores were missing at one point of measurement for at most 4.2% of students; mental ability scores (assessed at the first point of measurement only) were missing for 1.2% of students. The maximum percentage of missing data on the student questionnaires was 5.2%. In the teacher survey, 10.0% of teachers did not report on the disciplinary climate in the PISA class; 12.8% of teachers did not submit test and examination papers. In the following, we use the full information maximum likelihood algorithm implemented in M*plus*, which estimates the missing values using the full information of the covariance matrices at individual and class level under the "missing at random" assumption (Muthén & Muthén, 2004).

Results

This section begins with a descriptive overview, after which we discuss differences between teachers with different training backgrounds. We then present the mediation models investigating possible effects of teachers' CK and PCK on instructional features and student progress, focusing on the differential effects of the two knowledge components. Finally, we address the moderation hypothesis, which predicts that the effects of PCK are stronger in lower-achieving classes.

Descriptive Findings

The sample of 194 classes comprised 80 academic track classes and 114 classes distributed fairly equally across the nonacademic tracks. Given their structural similarities, the nonacademic tracks were collapsed into a single category for the following analyses. As shown in Table 1, the academic track is selective not only in terms of mental ability, prior knowledge of mathematics, and reading literacy but also in terms of SES, parental education, and

Table 1
Descriptive Findings at Student Level (*N* = 4,353)

	Total		Nonacademic Tracks		Academic Track		Δ_{tracks}	
Variables	*M*	*SD*	*M*	*SD*	*M*	*SD*	*t*	*p*
Mathematics achievement, end of Grade 10	0.05	0.98	–0.31	0.90	0.53	0.86	–30.3	.000
Mathematical literacy, end of Grade 9	0.04	0.97	–0.32	0.92	0.53	0.87	–31.6	.000
Reading literacy, end of Grade 9	0.06	0.96	–0.25	0.92	0.47	0.85	–19.3	.000
Mental ability, end of Grade 9	0.05	0.96	–0.22	0.96	0.43	0.81	–23.4	.000
Age, years	15.7	0.66	15.8	0.71	15.6	0.67	10.4	.000
Socioeconomic status[a]	53.4	16.0	48.6	14.6	59.9	15.4	–24.4	.000
	%		%		%		χ^2, (df = 1)	*p*
Parents with university degree	32.3		19.1		50.7		40.9	.000
Immigrant background	19.3		21.4		16.6		5.7	.000

[a]As operationalized by the International Socio-Economic Index. Specifically, the analysis used the higher-status occupation between the two parents.

ethnic background. The higher mean age of the students in the nonacademic tracks points to higher levels of grade retention.

Between-track differences were also found at the class level (see Table 2). The cognitive and curricular levels of the tasks set were somewhat lower in the classes of the nonacademic track. The level of cognitive challenge was low overall, however. Most of the tasks set were purely technical; few required mathematical reasoning and argumentation or conceptual flexibility. No consistent between-track differences were found for individual learning support or classroom management. Teachers of the academic and nonacademic tracks did not differ statistically significantly in terms of their age (*M* = 48 years), teaching experience (*M* = 22 years), gender distribution (52% male), or high school grade point average.

Teacher Training Program Attended and Subject Matter Knowledge

As shown in Table 3, teachers' CK and PCK scores proved to be highly dependent on the type of training program they had attended (see Krauss, Baumert, et al., 2008). As predicted, teachers certified to teach in the academic track (Certification Type 1) had much higher CK scores than did

Table 2
Descriptive Findings at Class Level (n = 194) and Teacher Level (n = 181)

	Total		Nonacademic Tracks		Academic Track		Δ_{tracks}	
Constructs and Variables	M^a	*SD*	M^a	*SD*	M^a	*SD*	*t*	*p*
Class Level								
Cognitive level of tasks								
Type of mathematical task[b]	1.58	0.25	1.57	0.24	1.59	0.26	–0.68	.50
Level of mathematical argumentation[c]	0.07	0.10	0.04	0.06	0.12	0.13	–5.50	.000
Innermathematical translation[c]	0.38	0.30	0.38	0.34	0.38	0.30	0.02	.98
Curricular level of tasks								
Alignment to grade 10 curriculum[b]	2.72	0.19	2.68	0.20	2.78	0.15	–3.38	.000
Individual learning support								
Adaptive explanations[d]	2.82	0.43	2.84	0.43	2.78	0.44	0.95	.34
Constructive response to errors[d]	2.94	0.44	2.91	0.44	2.97	0.43	–0.84	.40
Patience[d]	2.76	0.53	2.76	0.50	2.75	0.58	0.07	.95
Adaptive pacing[d]	2.20	0.44	2.20	0.44	2.21	0.44	–0.21	.84
Respectful treatment of students[d]	3.22	0.49	3.16	0.50	3.31	0.47	–2.10	.04
Caring ethos[d]	2.66	0.50	2.69	0.50	2.61	0.48	1.20	.23
Quality of classroom management								
Prevention of disruption[d]	2.53	0.62	2.52	0.61	2.56	0.63	–0.47	.64
Effective use of time[d]	2.66	0.58	2.66	0.56	2.65	0.60	0.21	.84
Teacher Level								
Age (in years)	48.15	8.0	48.51	7.8	47.46	8.4	0.79	.43
Gender (male)	52.1%		49.5%		57.1%		$\chi^2 = 0.85$	.36
Years of service	22.1	9.4	22.7	9.5	20.9	9.3	1.13	.26
Grade point average[e]	*Mdn* = 2.6		*Mdn* = 2.8		*Mdn* = 2.4		5.66	.35

[a]Unless noted otherwise.
[b]1 = low, 3 = high.
[c]0 = low, 3 = high.
[d]1 = low, 4 = high.
[e]1 = highest, 6 = lowest.

Table 3

Teachers' Content Knowledge and Pedagogical Content Knowledge by Type of Certification

	Total	Type 1: Academic	Type 2: Nonacademic	Type 3: Integrated[a]	$\Delta_{(\text{Type 1} - \text{Type 2})}$	$\Delta_{(\text{Type 1} - \text{Type 3})}$	$\Delta_{(\text{Type 2} - \text{Type 3})}$
Variables	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*t* (*p*)	*t* (*p*)	*t* (*p*)
CK	–0.139 (0.99)	0.737 (0.91)	–0.524 (0.82)	–0.446 (0.73)	7.537 (.000)	6.099 (.000)	–0.465 (.643)
PCK	–0.027 (0.99)	0.427 (1.10)	–0.007 (0.83)	–0.596 (0.77)	2.301 (.023)	4.442 (.000)	3.329 (.001)
PCK|CK[b]	–0.083	–0.097	0.256	–0.407	–1.909 (.058)	1.531 (.128)	4.064 (.000)

Note. Excluding the 14 teachers who completed a 5-year training program in the former German Democratic Republic. CK = content knowledge; PCK = pedagogical content knowledge.

[a]Training in former German Democratic Republic.

[b]PCK estimated after controlling for CK; PCK estimated at CK = –0.139.

teachers certified for nonacademic tracks (Certification Type 2) or teachers who attended an integrated training program in the former German Democratic Republic (Certification Type 3). The mean difference between teachers of Type 1 versus Type 2 was d = 1.26 *SD*; between teachers of Type 1 versus Type 3, d = 1.18 *SD*. Teachers of Type 2 versus Type 3 did not differ statistically significantly in their CK scores. The differences remained practically unchanged when selective intake was controlled in terms of the final high school grade point average (see Krauss, Brunner, et al., 2008). The dramatic differences of more than one standard deviation in CK scores may be attributable to the higher requirements placed on teacher candidates in the academic program and/or to the potentially stricter demands of mathematics departments. These two effects are confounded and cannot be separated in our study.

The teachers trained for the academic track also outscored their colleagues on PCK, even though the training requirements for teaching methods courses tend to be higher, if at all, in Type 2 programs and the courses were given by the same staff. As predicted, however, the mean difference in PCK between Certification Type 1 and Type 2 was much smaller than for CK (d = 1.26 *SD*), at d = 0.43 *SD*. When CK was controlled, Type 2 teachers in fact outscored Type 1 teachers on PCK (d = 0.35 *SD*). The PCK of Type 3 teachers was much lower than that of Type 1 and Type 2 teachers (d = 1.02 and d = 0.60, respectively), although the CK level of Type 2 and Type 3 teachers did not differ.

To test whether the differences in CK and PCK leveled out over the teaching career, we specified interactions between certification type and years of service. No significant interactions emerged, indicating that the differences persisted over the entire teaching career.

As outlined above, teachers in Germany are generally assigned to schools of the different tracks on the basis of their certification. As a result, there are marked differences in the CK and PCK of the teachers of the different tracks (see Table 4). In the United States, it is a matter of grave concern that the least qualified teachers work at low SES schools attended by the lowest-achieving students (Nye et al., 2004). In Germany, the correspondence between teacher training and the tracking system produces similar effects. The implications of this mechanism are discussed below.

Mathematical Knowledge of Teachers, Classroom Instruction, and Student Progress

In this section, we test the hypothesis that it is PCK and not CK that has the decisive impact on key aspects of instructional quality—namely, cognitive activation and individual learning support—and thus on student progress. We first present the results for the PCK model and then the CK model before comparing the two.

Table 4
Teachers' Content Knowledge and Pedagogical Content Knowledge by Track

	Total		Nonacademic Tracks		Academic Track		Δ_{tracks}	
Variables	*M*	*SD*	*M*	*SD*	*M*	*SD*	*t*	*p*
CK	–0.11	1.00	–0.58	0.84	0.73	0.68	–10.2	.000
PCK								
Tasks	–0.04	0.97	–0.24	0.89	0.31	1.00	–3.46	.001
Students	–0.02	1.00	–0.24	0.98	0.45	0.86	4.68	.000
Instruction	–0.03	0.97	–0.31	0.93	0.48	0.85	–5.37	.000
Overall	–0.05	0.98	–0.33	0.96	0.49	0.76	–5.65	.000

Note. CK = content knowledge; PCK = pedagogical content knowledge.

A series of multilevel structural equation models with latent variables was specified to test the mediation model presented in Figure 1. The measurement model for the latent variables in the full mediation model is shown in Table A2 of the appendix. All manifest indicators made a substantial contribution to defining the respective latent construct. The estimates for the structural parameters of the fitted models are summarized in Table 5.

In a first step, the variance in mathematics achievement was decomposed into within- and between-class components (unconditional model). The results showed that 54.5% of the variation in achievement was within classes and that 45.5% was between classes, highlighting the effects of early tracking in the German school system. When the between-class variance was partitioned into a between-track component (academic versus nonacademic) and a between-class within-track component, 23.5% and 22.0%, respectively, of the variance was explained.

In a second step, we specified the individual model (Model 1, Table 5), which we propose to reflect the mechanism of student assignment to different classes and teachers. We estimated a random intercept model with 10 achievement predictors, all of which were assessed at the end of Grade 9. As shown in Table 5 (Model 1), the decisive control variables at the individual level were prior knowledge and mental ability. The most important predictor was that of mathematical knowledge at the end of Grade 9 (β = .49), followed by mental ability (β = .24), and reading literacy (β = .21). Consistent with our assignment model, social background, parental education, and immigration status proved to be less important. The individual model explained a total of 64% of the variance in mathematics achievement at the end of Grade 10 *within* classes.

Because in Germany allocation to classes is highly dependent on student achievement and social background, the variance between

Table 5

Predicting Mathematics Achievement at the End of Grade 10 by Dimensions of Instructional Quality, Content Knowledge, and Pedagogical Content Knowledge, Controlling for Individual Selection Variables

	Individual	Dimensions of Instructional Quality		Black Box (PCK)		Mediation (PCK)	Black Box (CK)	Mediation (CK)
Predictor/Parameter	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Individual Level[a]								
Mathematics achievement, end of Grade 10 on:								
Prior knowledge of mathematics, end of Grade 9	**.49**	**.49**	**.49**	**.49**	**.49**	**.48**	**.49**	**.49**
Reading literacy, end of Grade 9	**.21**	**.21**	**.20**	**.21**	**.20**	**.20**	**.20**	**.20**
Mental ability, end of Grade 9	**.24**	**.23**	**.24**	**.24**	**.24**	**.24**	**.24**	**.24**
Socioeconomic status[b]	.01	.00	.00	.00	.00	–.01	–.01	–.01
Parental education (6 dummies)								
Hauptschule, no apprenticeship (0/1)	–.02	–.01	–.02	–.02	–.02	–.02	–.02	–.02
Hauptschule and apprenticeship (0/1)	–.02	–.02	–.02	–.02	–.02	–.02	–.02	–.02
Realschule and apprenticeship (0/1)	**–.03**	–.03	–.03	–.03	–.03	–.03	–.03	–.03
Realschule and professional training (reference)								
Gymnasium (0/1)	–.02	–.02	–.02	–.02	–.02	–.02	–.02	–.02
University (0/1)	.03	.04	.02	.03	.02	.02	.02	.02
Immigration status (0/1)	–.03	–.03	–.03	–.03	–.03	–.03	–.03	–.03
R^2	**.64**	**.64**	**.62**	**.62**	**.62**	**.62**	**.62**	**.62**

(continued)

Table 5 **(continued)**

	Individual	Dimensions of Instructional Quality		Black Box (PCK)		Mediation (PCK)	Black Box (CK)	Mediation (CK)
Predictor/Parameter	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Class Level								
Mathematics achievement, end of Grade 10 on:								
Track (nonacademic/academic)			**.58**		**.42**	**.56**	**.42**	**.60**
PCK (CK)				**.62**	**.42**		**.30**	
Cognitive level of tasks		**.32**	**.32**			**.32**		**.31**
Curricular level of tasks		**.30**	**.17**			**.17**		**.18**
Individual learning support		.08	.11			.10		.10
Effective classroom management		**.33**	**.30**			**.31**		**.32**
Cognitive level (MV1) on PCK (CK)						**.24**		.01
Curricular level (MV2) on PCK (CK)						**.33**		**.32**
Individual learning support (MV4) on PCK (CK)						**.26**		–.06
Effective classroom management (MV3) on PCK (CK)						.14		–.01
R^2		**.37**	**.68**	**.39**	**.54**	**.69**	**.44**	**.65**
χ^2	25.4	787	819	31.8	33.6	882	31.5	874
df	9	190	200	15	18	253	14	240
p	.003	.00	.00	.007	.07	.00	.005	.00
Comparative fit index	.99	.97	.97	.99	.99	.96	.99	.96
Root mean square error of approximation	.02	.03	.03	.02	.01	.02	.02	.03
Standardized root mean square residual between	.001	.04	.04	.06	.01	.08	.01	.09
Standardized root mean square residual within	.005	.03	.03	.005	.004	.02	.004	.02

Note. Multilevel structural equation models with latent variables, completely standardized solutions; coefficients in bold are significant at least at the 5% level. CK = content knowledge; PCK = pedagogical content knowledge; MV = mediator variable.

[a]All individual variables centered at the grand mean.

[b]As operationalized by the International Socio-Economic Index. Specifically, the analysis used the higher-status occupation between the two parents.

classes decreased dramatically when these covariates were controlled. The residual variance between school classes was only $\rho = .046$. In other words, a maximum of 4.6% of the variance in achievement at the individual level can be explained by different treatment at class level. The magnitude of this potential effect is comparable with findings from other studies (Hill et al., 2005; Lanahan, McGrath, McLaughlin, Burian-Fitzgerald, & Salganik, 2005; Nye et al., 2004).

In the next step, the four core dimensions of instructional quality were entered in the model (Table 5, Model 2). The predictors at class level were as follows: cognitive level of tasks, curricular level of tasks, individual learning support, and quality of classroom management. With the exception of individual learning support and classroom management, the correlation of which was $r_{\text{latent}} = .41$, the latent constructs of instructional quality were orthogonal. Cognitive level of tasks, curricular level of tasks, and effective classroom management proved to be significant predictors of mathematics achievement at the end of Grade 10. Contrary to our expectations, however, individual learning support was not found to have a specific effect on mathematics achievement over and above the joint effect with classroom management (the zero-order correlation between support and achievement was $r_{latent} = .22$). The four latent predictors explained 37% of the variance between classes.

Model 3 tested whether the model also holds when controlling for the classes' track membership. The high coefficient of $\beta = .58$ for school type (nonacademic versus academic track) indicates that the tracks constitute differential developmental environments. In principle, however, the instructional model also holds within school types. Only the curricular level of tasks was found to be confounded with track membership, which caused the standardized regression coefficient to drop from $\beta = .30$ to $\beta = .17$. Model 3 explained 68% of the variance in achievement between classes.

Models 4 and 5 tested whether the PCK of Grade 10 mathematics teachers is in fact relevant to students' achievement. These black-box models test the direct effects of PCK on mathematics achievement at the end of Grade 10, with and without control for track membership. The findings are clear. The standardized regression coefficient for PCK in Model 4—without control for track membership—was $\beta = .62$. In other words, 39% of the variance in achievement between classes was explained solely by the latent variable of PCK. Teachers' domain-specific instructional knowledge thus seems to be of key significance for student progress in mathematics. The relationship between PCK and mathematics achievement was linear. An additionally estimated quadratic term was insignificant (cf. Hill et al., 2004).

As shown earlier, track membership and PCK are confounded; as Model 5 shows, however, they can still be distinguished empirically (track membership was entered as a control variable in Model 5). Track membership and teachers' PCK each had considerable specific significance for students' learning gains (both $\beta = .42$). The explained variance between classes was $R^2 = .54$. The shared variance component was $R^2 = .23$; the track-specific variance component, $R^2 = .17$; and the PCK-specific variance component, $R^2 = .14$.

Model 6 tested the full mediation model, controlling for the track membership of the classes investigated. The parameter estimates of the relationship between instruction and achievement are comparable to those reported for Model 2. As expected, PCK seems to influence the cognitive level, curricular level, and learning support dimensions of instructional quality. The finding that PCK affects individual learning support in mathematics is particularly interesting as it shows that learning support seems to be dependent not only on a caring ethos but also on domain-specific knowledge. The independence of classroom management from PCK can be interpreted as an indicator for the discriminant validity of PCK. It is possible for classroom management to seem effective on the surface, even when levels of PCK and cognitive activation are low. The full mediation model explained 69% of the variance in achievement between classes.

What Counts: CK or PCK?

Previous findings have shown the substantial correlations between CK and PCK to increase as a function of the expertise of the teacher group (Krauss, Brunner, et al., 2008). These findings raise the urgent question of whether PCK or CK is decisive in the classroom or whether the two components of professional knowledge are interchangeable. Our theoretical assumption is that PCK is inconceivable without a substantial level of CK but that CK alone is not a sufficient basis for teachers to deliver cognitively activating instruction that, at the same time, provides individual support for students' learning.

To address this question, we also specified the black-box model (PCK predicting student achievement after controlling for track; Model 5 in Table 5) for CK (Model 7 in Table 5). Model fit was similar to that of Model 5. CK was less predictive of student progress than was PCK, however ($\beta = .30$ versus $\beta = .42$). At the same time, the explained between-class variance decreased from $R^2 = .54$ to $R^2 = .44$. PCK thus has greater power than CK to explain student progress.

To examine the implications of these findings for the instructional process, we also specified the full mediation model for CK (Model 8). When the parameters of the regression of the instructional variables on CK were freely estimated, the distinct effects of PCK and CK became apparent. CK was not found to affect the cognitive level of tasks and individual learning support; the coefficients were practically zero. It was only the curricular level of the tasks—that is, their curricular alignment—that increased with increasing levels of CK (β = .32). The observed differences between PCK and CK as predictors for instruction and student learning can also be statistically substantiated. To this end, we constrained the critical parameters of the regression of the instructional variables on CK in Model 8 to the standardized values estimated for PCK in Model 6. Under these conditions, the fit of Model 8 was significantly reduced: The difference in χ^2 at 4 degrees of freedom was 18.6 ($p < .05$); Akaike's information criterion increased from 129,102 to 129,118. Our findings thus confirm that it is PCK that has greater predictive power for student progress and is decisive for the quality of instruction. These results do not imply that CK has no direct influence on instructional features, however. In fact, teachers with higher CK scores are better able to align the material covered with the curriculum. But higher levels of CK have no direct impact either on the potential for cognitive activation or on the individual learning support that teachers are able to provide when learning difficulties occur. It is the level of PCK that is decisive in both these cases. Both PCK and CK vary independently of effective classroom management.

Effect Size of PCK

The mediation model specified for PCK explained 39% of the variance in achievement between classes *without* control for track membership (not reported in Table 5). The amount of variance explained was thus identical with the effect of PCK in the black-box model (Model 4 in Table 5). What are the practical implications of this finding? To facilitate interpretation, it is helpful to evaluate effect sizes based on students' general learning rates. The mean increase in mathematics across Grade 10 in our sample was d = 0.35. To transform the variance component attributable to teachers' PCK into an interpretable effect size, we chose a procedure based on Tymms' proposal (2004) for calculating effect sizes for continuous Level 2 predictors in multilevel models. This effect size, which is comparable with Cohen's d, can be calculated using the following formula:

$$\Delta = 2 \times B \times SD_{\text{predictor}}/\sigma_e$$

where B is the unstandardized regression coefficient in the multilevel model, $SD_{\text{predictor}}$ is the standard deviation of the predictor variable at the class level, and σ_e is the residual standard deviation at the student level. The resulting effect size describes the difference in the dependent variable between two classes that differ two standard deviations on the predictor variable. This gives—without control for track membership—a

PCK effect of d_{class} = 0.46 (SE = .09). In other words, two comparable Grade 10 classes whose mathematics teachers' PCK differed by two standard deviations would differ by d = 0.46 SD in their mean mathematics achievement at the end of the school year. Based on the average student learning rate of d = 0.35, completely comparable classes taught by teachers with PCK scores in the lower or upper quintile of the competence distribution can thus be expected—all things being equal—to show learning gains in the range of about $d \leq 0.15$ and $d \geq 0.55$, respectively. This effect size may be overestimated because no account is taken of track membership. When track membership is controlled (Model 5), the effect size for the specific PCK effect is d_{class} = 0.328 (SE = .10). This effect size may be underestimated because no account is taken of the confounded effect component (see above, R^2 = .227). In this case, classes taught by teachers with PCK scores in the lower or upper quintile of the competence distribution can be expected to show learning gains of $d \leq .21$ and $d \geq .49$, respectively. The true effect size lies somewhere between the two estimates and is thus substantial.

Moderating Effects of Track

To test the hypothesis that teachers' PCK is particularly important for the learning gains of weaker students, we also specified Model 4 (cf. Table 5) as a two-group model (not reported in the tables) in which model parameters were estimated separately for classes in the nonacademic tracks (low SES and low achievement) and for Gymnasium classes (high SES and high achievement). We tested the moderator effect by comparing the fit indices of the two-group model when the effect of PCK was freely estimated versus constrained to be equal. When freely estimated, the standardized regression coefficient of student achievement on PCK was β = .54 in the first group and β = .29 in the second group. These findings indicate that differences in teacher PCK have a greater impact on students in low SES low-achievement classes. The fit of the two-group model was excellent, χ^2 = 57.9, df = 46, p = .11; comparative fit index = .99; root mean square error of approximation = .011; standardized root mean square residual$_{between}$ = .03 and standardized root mean square residual$_{within}$ = .005. However, model fit was only minimally reduced when the regression coefficients were constrained to be equal. The difference in χ^2 at 1 degree of freedom was 2.0 and was not significant. The interaction between track and PCK did not reach the level of statistical significance.

Summary and Conclusions

Learning and instruction are domain specific. As Leinhardt (2001) has shown with reference to instructional explanations in history and mathematics, the structure and syntax of the subject affect instructional processes and necessitate specific teacher expertise, which can be acquired through formal training and reflected teaching experience (Ball et al., 2001; Grossman & Schoenfeld, 2005). In our study, we investigated the subject-specific knowledge of secondary school mathematics teachers, and our results confirmed the relevance of these forms of specific teacher expertise for high-quality teaching and student learning. We considered both CK and PCK as critical professional resources for teachers, each requiring specific attention during both teacher training and classroom teaching practice.

In contrast to Hill and colleagues (Hill et al., 2004; Hill et al., 2007), who conceptualized mathematical knowledge for teaching as an amalgam of the mathematical everyday knowledge that all educated adults should have, a purely mathematical understanding of topics typically taught at school, and mathematical knowledge relating directly to the instructional process (PCK), the COACTIV group has succeeded in distinguishing the CK and PCK of secondary mathematics teachers conceptually and empirically. In line with the findings of qualitative studies on teacher knowledge, the COACTIV group worked on the theoretical assumption that PCK as a specific form of mathematical knowledge is inconceivable without sufficient CK but that CK cannot substitute PCK. Unlike CK, PCK was expected to be manifested in the quality of the instructional process itself. This hypothesis was tested by means of model comparison, using hierarchical structural equation models with latent variables.

When selective intake to schools and classes was controlled at the individual level, PCK explained 39% of the between-class variance in achievement at the end of Grade 10. The effect sizes were substantial: If two learning groups comparable at the beginning of Grade 10 were taught by mathematics teachers whose PCK differed by two standard deviations, the groups' mean mathematics achievement would differ by d = 0.46 *SD* across all tracks or d = 0.33 *SD* within tracks by the end of the school year. This effect was fully mediated by the level of cognitive activation provided by the tasks set, instructional alignment with the Grade 10 curriculum, and individual learning support. In other words, PCK largely determines the cognitive structure of mathematical learning opportunities.

The mediation model does not apply to CK, or only to a very limited extent. Despite its high correlation with PCK, CK has lower predictive power for student progress. CK has a direct impact only on the alignment of tasks to the Grade 10 curriculum. No direct effects were found on the two key variables of instructional quality—namely, cognitive activation and individual learning support.

This does not imply that CK—defined as a conceptual understanding of the mathematical knowledge taught—is unimportant. As shown by the qualitative studies reviewed in our overview of the research literature above, CK defines the possible scope for the development of PCK and for the provision of instruction offering both cognitive activation and individual support. This study shows that both CK and PCK are largely dependent on the type of training program attended, with program-specific differences in CK of more than one standard deviation. The qualitative findings indicate that deficits in CK are to the detriment of PCK, limiting the scope for its development. Our findings suggest that it is not possible to offset this relationship by increasing the specific focus on PCK in teacher training.

These results provide broad, representative confirmation for findings on the structure and effects of domain-specific professional knowledge that have accumulated over the past two decades (Ball et al., 2001). Further, they extend on the findings reported by Hill et al. (2004) for elementary school teachers.

Our findings also allow some tentative conclusions to be drawn for the structure and design of teacher training programs. It seems that programs that compromise on subject matter training, with the result that teacher candidates develop only a limited mathematical understanding of the content covered at specific levels, have detrimental effects on PCK and consequently negative effects on instructional quality and student progress. Differences in CK that emerge during preservice training persist across the entire teaching career.

This does not imply that it is the best possible solution for mathematics teacher candidates to attend training programs largely identical to those provided for students majoring in mathematics, although such programs do seem to produce better results than the shorter and possibly more superficial training programs implemented at German schools of education. It is probably also possible to achieve sound understanding of the structure and syntax of the discipline without loss of mathematical rigor by reference to school-related topics. A challenge for future research on teacher training will be to examine whether and how this can best be achieved.

PCK also depends on the type of training program attended. Here again, deficits in training programs are not offset by practical on-the-job experience. There is much to suggest that training programs must achieve a balance between CK and PCK. Determining the nature of this balance is a further challenge for future research.

The ongoing discussion on the certification of "highly qualified" teachers in the United States focuses on subject matter knowledge (U.S. Department of Education, 2003, 2006). Critics are very concerned about this one-sided approach (Grossman, 2008; Liston, Borko, & Whitcomb, 2008; Smith & Gorard, 2007). Our findings show that it is, in principle, justified to increase the attention paid to teachers' subject matter knowledge. However, it is vital to specify from the outset exactly what is meant by subject matter knowledge—in terms of CK, PCK, and a balance between the two. It is clear that these knowledge components do not exhaust the spectrum of professional competence. Indeed, the substantial amount of unexplained variance in our structural equation model clearly shows that the choreography of teaching is not only dependent on CK and PCK. Further research is thus needed into the theoretical conceptualization and operationalization of generic pedagogical knowledge (e.g., Blömeke, Felbrich, & Müller, 2008).

The second conclusion to be drawn from our findings is a sociopolitical one relating to the equality of opportunities. In Germany, teacher candidates aspiring to teach at secondary level attend different training programs depending on the school track in which they intend to teach. Teachers graduating from these programs differ considerably in their mathematical knowledge. Because the allocation of teachers to school types is relatively strict, the professional competence of mathematics teachers in the academic versus nonacademic tracks differ accordingly, with serious implications for social equality.

Students attending the different tracks differ not only in their ability and achievement but also in their social and ethnic backgrounds. Consequently, weaker students from lower SES families and immigrant families tend to be taught by teachers who are less competent in terms of CK and PCK. This is one of the factors contributing to the extremely wide distribution in achievement and the serious social and ethnic disparities found in the United States and Germany as well at the end of compulsory schooling (Akiba, LeTendre, & Scribner, 2007; Baumert & Schümer, 2001). The unequal distribution of well-trained teachers across schools is a matter of great concern in the United States, where it is primarily the result of differences in the social structure of school districts (Darling-Hammond, 2006; Hill & Lubienski, 2007; Zumwalt & Craig, 2005). In Germany, it is caused largely by an interaction of the institutional structure of the education system and the structure of teacher training.

Debates on the improvement of teacher quality often prove controversial. In many cases, outside observers call the existing systems of teacher education into question and suggest that the teaching profession be opened to subject experts without specific pedagogical training. We hope that the empirical evidence provided by our study will inform these discussions. We see the key message of our study as follows: PCK—the area of knowledge relating specifically to the main activity of teachers, namely, communicating subject matter to students—makes the greatest contribution to explaining student progress. This knowledge cannot be picked up incidentally, but as our finding on different teacher-training programs show, it can be acquired in structured learning environments. One of the next great challenges for teacher research will be to determine how this knowledge can best be conveyed to both preservice and in-service teachers.

Appendix

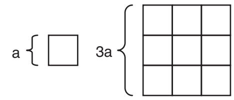
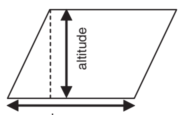
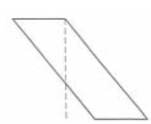
Knowledge Category (Subscale)	Sample Item	Sample Response (Scored as correct)
CK	Is it true that 0.999999 . . . = 1? Please give detailed reasons for your answer.	Let 0.999 . . . = a Then 10a = 9.99 . . . , hence, $\underbrace{10a - a}_{9a} = \underbrace{9.99\ldots - 0.999\ldots}_{9}$ Therefore a = 1; hence, the statement is true.
PCK: ***tasks***	How does the surface area of a square change when the side length is tripled? Show your reasoning. Please note down as many different ways of solving this problem (and different reasonings) as possible.	*Algebraic response* Area of original square: a^2 Area of new square is then $(3a)^2 = 9a^2$; i.e., 9 times the area of the original square. *Geometric response* Nine times the area of the original square: 
PCK: ***students***	The area of a parallelogram can be calculated by multiplying the length of its base by its altitude.  Please sketch an example of a parallelogram to which students might fail to apply this formula.	 *Note:* The crucial aspect to be covered in this teacher response is that students might run into problems if the foot of the altitude is outside a given parallelogram.
PCK: ***instruction***	A student says: I don't understand why $(-1) \times (-1) = 1$ Please outline as many different ways as possible of explaining this mathematical fact to your student.	The "permanence principle," although it does not prove the statement, can be used to illustrate the logic behind the multiplication of two negative numbers and thus foster conceptual understanding: $3 \times (-1) = -3$ $2 \times (-1) = -2$ $-1 \quad 1 \times (-1) = -1 \quad +1$ $0 \times (-1) = 0$ $(-1) \times (-1) = 1$ $(-2) \times (-1) = 2$

Figure A1. **Measures of pedagogical content knowledge and content knowledge in COACTIV: Sample items and item responses.**

Note. CK = content knowledge; PCK = pedagogical content knowledge.

Type of Mathematical Task	Curricular Level of Task	
	Low Knowledge of elementary mathematics: basic arithmetical operations and a knowledge of the basic geometry taught at elementary level or known from everyday life.	***High*** Knowledge of advanced junior high school mathematics: advanced procedures and concepts (e.g., quadratic equations, rudiments of conformal geometry)
Technical tasks Technical knowledge without any contextual anchoring	Calculate the following sum: $13.4 \text{ liters} + 3 \text{ dm}^3 = \_\_\_\_ \text{ dm}^3$	Simplify the following fraction: $\frac{4b^{-1}c^{-1}fj^2n^5}{2b^{-1}c^{-2}fj^2n^5} =$
Conceptual modeling tasks Inner- and extramathematical tasks that require modeling, and that emphasize conceptual reasoning in the procedural phase	Continue the following sequence: 1 4 9 16 25 ________	A sum of 4,000 is invested in an account that pays 4.5% interest per year. Express the relationship as a function.

Figure A2. **Examples of tasks of differing cognitive demands and curricular levels.**

Table A1
Dimensions of Instructional Quality: Sample Items and Reliabilities of Indicators

Latent Constructs and Indicators	Items (n)	Sample Items	ICC_1	ICC_2[a]
Individual learning support		*Our mathematics teacher . . .*		
Adaptive explanations	4	Gives good examples that make math problems easy to understand.	.34	.86
Constructive response to errors	3	Doesn't mind if someone makes a mistake in the lesson.	.32	.85
Patience	3	Stays patient even if he/she has to explain things several times.	.37	.86
Adaptive pacing	4	Often doesn't really discuss our problems because we have so much material to get through.[b]	.29	.83

(continued)

Table A1 **(continued)**

Latent Constructs and Indicators	Items (n)	Sample Items	ICC_1	ICC_2[a]
Respectful treatment of students	3	Sometimes treats students in a hurtful way.[b]	.31	.85
Caring ethos	3	Takes time to listen whenever a student wants to talk to him/her.	.35	.86
Effective classroom management				
Disciplinary climate (teacher perception)	8	My lessons in this class are very often disrupted.[b]	—	—
Prevention of disruption	3	Our mathematics lessons are very often disrupted.[b]	.44	.90
Effective use of time	3	A lot of time gets wasted in mathematics lessons.[b]	.41	.89

Note. ICC = intraclass correlation coefficient. The ICC_1 indicates the proportion of total variance that can be attributed to between-class differences. It is a measure of the reliability of an *individual* student's judgment. The ICC_2 indicates the reliability of a class-mean judgment and is calculated by applying the Spearman-Brown formula to ICC_1. Dashes (—) indicate that measure was not administered at individual level, thus ICCs cannot be computed.
[a]On average, 12 student ratings per class.

Table A2
Measurement Models for the Latent Constructs of Instruction and Achievement

	Factor Loadings	
Factors and Indicators	Individual Level	Class Level
Mathematical competence, end of Grade 10		
Test Part A (ST)	.74	.90
Test Part B (ST)	.75	.90
Pedagogical content knowledge		
Tasks (TT)	—	.53
Students (TT)	—	.67
Instruction (TT)	—	.78
Content knowledge		
Test Part A (TT)	—	.73
Test Part B (TT)	—	.83
Cognitive level of tasks		
Type of mathematical task (E)	—	.83
Level of mathematical argumentation (E)	—	.47
Innermathematical translation (E)	—	.70
Curricular level of tasks		

(continued)

Table A2 **(continued)**

Factors and Indicators	Factor Loadings	
	Individual Level	Class Level
Alignment to Grade 10 curriculum (E)	—	1.0
Individual and respectful learning support		
Adaptive explanations (S)	.78	.92
Constructive response to errors (S)	.77	.95
Patience (S)	.78	.96
Adaptive pacing (S)	.55	.86
Respectful treatment of students (S)	.49	.80
Caring ethos (S)	.68	.93
Effective classroom management		
Disciplinary climate (T)	—	.56
Prevention of disruption (S)	.79	.98
Effective use of time (S)	.80	.98

Note. E = tasks implemented in written examinations and tests; S = student reports; T = teacher reports; ST = student test; TT = teacher test. Dashes (—) indicate that measure was not administered at individual level, thus ICCs cannot be computed.

Notes

The research reported in this article is based on data from the COACTIV study (COACTIV: Professional Competence of Teachers, **Co**gnitively **Activ**ating Instruction, and the Development of Students' Mathematical Literacy), which was funded by the German Research Foundation (DFG; BA 1461/2-2). We thank Oliver Lüdtke for his comments and advice and Susannah Goss for translation and language editing.

[1]COACTIV: Professional Competence of Teachers, **Co**gnitively **Activ**ating Instruction, and the Development of Students' Mathematical Literacy.

[2]Our sample also includes a small number of teachers (n = 14) who completed a 5-year training program in the former German Democratic Republic. These teachers were not included in our analyses of content knowledge and pedagogical content knowledge.

[3]One question that remains open is whether the self-related cognitions and motivational orientations of students in different tracks show differential developmental trajectories after the transition to secondary school, as has been found for systems implementing *within*-school tracking (Gamoran, Nystrand, Berends, & LePore, 1995). In this case, the classes in our sample would differ in achievement-related variables above and beyond those included in the assignment model. We therefore ran additional analyses for mathematical self-concept and mathematical interest. Findings showed that students in the academic and nonacademic tracks did not differ in these characteristics at the end of Grade 9. The schools of the different tracks evidently constitute separate frames of reference for the social comparisons that regulate students' self-concepts and interests (Marsh, Köller, & Baumert, 2001). As such, there was no need to include motivational variables in the assignment model.

References

Akiba, M., LeTendre, G. K., & Scribner, J. P. (2007). Teacher quality, opportunity gap, and national achievement in 46 countries. *Educational Researcher, 36*(7), 369–387.

American Council on Education. (1999). *To touch the future: Transforming the way teachers are taught*. Washington, DC: Author.

Attewell, P., & Domina, T. (2008). Raising the bar: Curricular intensity and academic performance. *Educational Evaluation and Policy Analysis*, *30*(1), 51–71.

Ball, D. L. (1990). The mathematical understandings that prospective teachers bring to teacher education. *The Elementary School Journal*, *90*(4), 449–466.

Ball, D. L. (1991). Research on teaching mathematics: Making subject-matter knowledge part of the equation. In J. Brophy (Ed.), *Advances in research on teaching: Vol. 2. Teachers' knowledge of subject matter as it relates to their teaching practice* (pp. 1–48). Greenwich: JAI Press.

Ball, D. L., & Bass, H. (2003). Toward a practice-based theory of mathematical knowledge for teaching. In B. Davis & E. Simmt (Eds.), *Proceedings of the 2002 annual meeting of the Canadian Mathematics Education Study Group* (pp. 3–14). Edmonton, Alberta, Canada: Canadian Mathematics Education Study Group/ Groupe Canadien d'étude En Didactique des Mathématiques.

Ball, D. L., Lubienski, S. T., & Mewborn, D. S. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 433–456). Washington, DC: American Educational Research Association.

Ballou, D., & Podgursky, M. (2000). Reforming teacher preparation and licensing: What is the evidence? *Teachers College Record*, *102*(1), 5–27.

Baumert, J., & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb [Family background, educational participation, and competency acquisition]. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, et al. (Eds.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (pp. 323–407). Opladen, Germany: Leske + Budrich.

Berliner, C. D. (1994). Expertise: The wonder of exemplary performances. In J. N. Mangieri & C. C. Block (Eds.), *Creating powerful thinking in teachers and students* (pp. 161–186). Fort Worth, TX: Holt, Rinehart & Winston.

Blömeke, S., Felbrich, A., & Müller, C. (2008). Messung des erziehungswissenschaftlichen Wissens angehender Lehrkräfte [Assessment of student teachers' pedagogical knowledge]. In S. Blömeke, G. Kaiser & R. Lehmann (Eds.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer—Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare: Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung* (pp. 171–194). Münster, Germany: Waxmann.

Blömeke, S., Kaiser, G., & Lehmann, R. (2008). *Professionelle Kompetenz angehender Lehrerinnen und Lehrer—Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare: Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung* [Professional competence of student teachers—Knowledge, beliefs, and learning opportunities of mathematics teacher candidates: Initial findings on the effectiveness of teacher training]. Münster, Germany: Waxmann.

Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models*. Newbury Park, CA: Sage.

Borko, H., Eisenhart, M., Brown, C., Underhill, R., Jones, D., & Agard, P. (1992). Learning to teach hard mathematics: Do novice teachers and their instructors give up too easily? *Journal for Research in Mathematics Education*, *23*(3), 194–222.

Borko, H., & Livingston, C. (1989). Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal*, *26*(4), 473–498.

Bransford, J., Darling-Hammond, L., & LePage, P. (2005). Introduction. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world* (pp. 1–39). San Francisco: Jossey-Bass.

Bransford, J. D., Derry, S. J., Berliner, C. D., & Hammerness, K. (2005). Theories of learning and their roles in teaching. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world* (pp. 40–87). San Francisco: Jossey-Bass.

Brennan, R. L. (2001). *Generalizability theory*. New York: Springer-Verlag.

Brophy, J. (2000). *Teaching*. Brussels, Belgium: International Academy of Education.

Carstensen, C. H., Knoll, S., Rost, J., & Prenzel, M. (2004). Technische Grundlagen [Technical foundations]. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, et al. (Eds.), *PISA 2003: Der Bildungsstand der Jugendlichen in Deutschland—Ergebnisse des zweiten internationalen Vergleichs* (pp. 371–387). Münster, Germany: Waxmann.

Cochran-Smith, M., & Zeichner, K. M. (2005). *Studying teacher education: The report of the AERA Panel on Research and Teacher Education*. Mahwah, NJ: Erlbaum.

D'Agostino, J. V., & Powers, S. J. (2009). Predicting teacher performance with test scores and grade point average: A meta-analysis. *American Educational Research Journal*, *46*(1), 146–182.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives*, *8*(1), 1–46.

Darling-Hammond, L. (2006). No Child Left Behind and high school reform. *Harvard Educational Review*, *76*(4), 642–667.

De Corte, E., Greer, B., & Verschaffel, L. (1996). Mathematics teaching and learning. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 491–549). Mahwah, NJ: Erlbaum.

Deng, Z. (2007). Knowing the subject matter of a secondary school science subject. *Journal of Curriculum Studies*, *39*(5), 503–535.

Ditton, H. (2007). *Kompetenzaufbau und Laufbahnen im Schulsystem: Ergebnisse einer Längsschnittuntersuchung an Grundschulen* [Competence building and careers in the education system: Results of a longitudinal study in elementary schools]. Münster, Germany: Waxmann.

Dubberke, T., Kunter, M., McElvany, N., Brunner, M., & Baumert, J. (2008). Lerntheoretische Überzeugungen von Mathematiklehrkräften: Einflüsse auf die Unterrichtsgestaltung und den Lernerfolg von Schülerinnen und Schülern [Mathematics teachers' beliefs and their impact on instructional quality and student achievement]. *Zeitschrift für Pädagogische Psychologie*, *22*(3–4), 193–206.

Eisenhart, M., Borko, H., Underhill, R., Brown, C., Jones, D., & Agard, P. (1993). Conceptual knowledge fall through the cracks: Complexities of learning to teach mathematics for understanding. *Journal for Research in Mathematics Education*, *24*(1), 8–40.

Even, R. (1993). Subject-matter knowledge and pedagogical content knowledge: Prospective secondary teachers and the function concept. *Journal for Research in Mathematics Education*, *24*(2), 94–116.

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education*, *27*(4), 403–434.

Fennema, E., & Franke, M. L. (1992). Teachers' knowledge and its impact. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 147–164). New York: Macmillan.

Floden, R. E., & Meniketti, M. (2005). Research on the effects of coursework in the arts and sciences and in the foundations of education. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying teacher education* (pp. 261–308). Mahwah, NJ: Erlbaum.

Gamoran, A., Nystrand, M., Berends, M., & LePore, P. C. (1995). An organizational analysis of the effects of ability grouping. *American Educational Research Journal*, *32*(4), 687–715.

Ganzeboom, H. B. G., & Treiman, D. J. (2003). Three internationally standardised measures for comparative research on occupational status. In J.H. P. Hoffmeyer-Zlotnik & C. Wolf (Eds.), *Advances in cross-national comparison: A European working book for demographic and socio-economic variables* (pp. 159–193). New York: Kluwer Academic/Plenum.

Goe, L. (2007). *The link between teacher quality and student outcomes: A research synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality.

Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, *22*(2), 129–145.

Goodson, I. F., Anstead, C. J., & Mangan, J. M. (1998). *Subject knowledge: Readings for the study of school subjects*. London: Falmer Press.

Goodson, I. F., & Marsh, C. (1996). *Studying school subjects*. London: Falmer Press.

Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 15–46). Mahwah, NJ: Erlbaum.

Grossman, P. L. (1995). Teachers' knowledge. In L. W. Anderson (Ed.), *International encyclopedia of teaching and teacher education* (*Vol. 2*, pp. 20–24). Oxford, UK: Pergamon Press.

Grossman, P. L. (2008). Responding to our critics: From crisis to opportunity in research on teacher education. *Journal of Teacher Education*, *59*(1), 10–23.

Grossman, P. L., & McDonald, M. (2008). Back to the future: Directions for research in teaching and teacher education. *American Educational Research Journal*, *45*(1), 184–205.

Grossman, P. L., & Schoenfeld, A. (2005). Teaching subject matter. In L. Darling-Hammond, J. Bransford, P. LePage, K. Hammerness & H. Duffy (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 201–231). San Francisco: Jossey-Bass.

Harbison, R. W., & Hanushek, E. A. (1992). *Educational performance of the poor: Lessons from rural northeast Brazil*. Oxford, UK: Oxford University Press.

Heaton, R. M. (2000). *Teaching mathematics to the new standards: Relearning the dance*. New York: Teachers College Press.

Heller, K., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision (KFT 4–12 + R)* [Cognitive Abilities Test for Grades 4 to 12, revision]. Göttingen, Germany: Hogrefe.

Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts* [Instructional quality and teacher professionality: Diagnosis, evaluation, and enhancement of instruction]. Seelze-Velber, Germany: Kallmeyer.

Hiebert, J., Morris, A. K., Berk, D., & Jansen, A. (2007). Preparing teachers to learn from teaching. *Journal of Teacher Education*, *58*(1), 47–61.

Hiebert, J., Stigler, J. W., Jacobs, J. K., Givvin, K. B., Garnier, H., Smith, M., et al. (2005). Mathematics teaching in the United States today (and tomorrow): Results from the TIMSS 1999 Video Study. *Educational Evaluation and Policy Analysis*, *27*(2), 111–132.

Hill, H. C. (2007). Mathematical knowledge of middle school teachers: Implications for the No Child Left Behind policy initiative. *Educational Evaluation and Policy Analysis*, *29*(2), 95–114.

Hill, H. C., Ball, D. L., Blunk, M., Goffney, I. M., & Rowan, B. (2007). Validating the ecological assumption: The relationship of measure scores to classroom teaching and student learning. *Measurement: Interdisciplinary Research and Perspectives*, *5*(2–3), 107–117.

Hill, H. C., & Lubienski, S. T. (2007). Teachers' mathematics knowledge for teaching and school context: A study of California teachers. *Educational Policy*, *21*(5), 747–768.

Hill, H. C., Rowan, B., & Ball, D. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*(2), 371–406.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal*, *105*(1), 11–30.

Jordan, A., Ross, N., Krauss, S., Baumert, J., Blum, W., Neubrand, M., et al. (2006). *Klassifikationsschema für Mathematikaufgaben: Dokumentation der Aufgabenkategorisierung im COACTIV-Projekt* [Classification scheme for mathematical tasks: Documentation of the task categorization in the COACTIV study]. Berlin: Max-Planck-Institut für Bildungsforschung (Materialien aus der Bildungsforschung 81).

Kahan, J., Cooper, D., & Bethea, K. (2003). The role of mathematics teachers' content knowledge in their teaching: A framework for research applied to a study of student teachers. *Journal of Mathematics Teacher Education*, *6*(3), 223–252.

Krauss, S., Baumert, J., & Blum, W. (2008). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: Validation of the COACTIV constructs. *The International Journal on Mathematics Education*, *40*(5), 873–892.

Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., et al. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology*, *100*(3), 716–725.

Kunter, M., & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, *9*(3), 231–251.

Kunter, M., Dubberke, T., Baumert, J., Blum, W., Brunner, M., Jordan, A., et al. (2006). Mathematikunterricht in den PISA-Klassen 2004: Rahmenbedingungen, Formen und Lehr-Lernprozesse [Mathematics education in the PISA classes 2004: Basic conditions, forms, and teaching and learning processes]. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, et al. (Eds.), *PISA 2003: Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (pp. 161–194). Münster, Germany: Waxmann.

Kunter, M., Klusmann, U., & Baumert, J. (2009). Professionelle Kompetenz von Mathematiklehrkräften: Das COACTIV-Modell [Teachers' professional competence: The COACTIV model]. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Eds.), *Lehrprofessionalität—Bedingungen, Genese, Wirkungen und ihre Messung* (pp. 153–164). Weinheim, Germany: Beltz.

Kunter, M., Klusmann, U., Dubberke, T., Baumert, J., Blum, W., Brunner, M., et al. (2007). Linking aspects of teacher competence to their instruction: Results from the COACTIV project. In M. Prenzel (Ed.), *Studies on the educational quality of schools: The final report on the DFG priority programme* (pp. 32–52). Münster, Germany: Waxmann.

Lanahan, L., McGrath, D. J., McLaughlin, M., Burian-Fitzgerald, M., & Salganik, L. (2005). *Fundamental problems in the measurement of instructional processes: Estimating reasonable effect sizes and conceptualizing what is important to measure*. Washington, DC: American Institutes for Research.

Leinhardt, G. (2001). Instructional explanations: A commonplace for teaching and location for contrast. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 333–357). Washington, DC: American Educational Research Association.

Liston, D., Borko, H., & Whitcomb, J. (2008). The teacher educator's role in enhancing teacher quality. *Journal of Teacher Education*, *59*(2), 111–116.

Lüdtke, O., Trautwein, U., Kunter, M., & Baumert, J. (2006). Reliability and agreement of student ratings of the classroom environment: A reanalysis of TIMSS data. *Learning Environments Research*, *9*(3), 215–230.

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Hillsdale, NJ: Erlbaum.

Marsh, H. W., Köller, O., & Baumert, J. (2001). Reunification of East and West German school systems: Longitudinal multilevel modeling study of the big-fish-little-pond effect on academic self-concept. *American Educational Research Journal*, *38*(3), 321–350.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist*, *59*(1), 14–19.

Mewborn, D. (2003). Teachers, teaching, and their professional development. In J. Kilpatrick, W. G. Martin & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 45–52). Reston, VA: National Council of Teachers of Mathematics.

Mitchell, R., & Barth, P. (1999). Not good enough: A content analysis of teacher licensing examinations: How teacher licensing tests fall short. *Thinking K–16*, *3*(1), 1–20.

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, *13*(2), 125–145.

Mullens, J. E., Murnane, R. J., & Willett, J. B. (1996). The contribution of training and subject matter knowledge to teaching effectiveness: A multilevel analysis of longitudinal evidence from Belize. *Comparative Education Review*, *40*(2), 139–157.

Munby, H., Russell, T., & Martin, A. K. (2001). Teachers' knowledge and how it develops. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 877–904). Washington, DC: American Educational Research Association.

Muthén, L. K., & Muthén, B. O. (2004). *Mplus statistical analysis with latent variables: User's guide* (3rd ed.). Los Angeles: Muthén & Muthén.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education.

Nye, B., Hedges, L. V., & Konstantopoulos, S. (2000). Effects of small classes on academic achievement: The results of the Tennessee class size experiment. *American Educational Research Journal*, *37*(1), 123–151.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*(3), 237–257.

Organisation for Economic Co-operation and Development. (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris: Author.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, *74*(4), 525–556.

Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research*, *63*(2), 167–199.

Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., et al. (2006). *PISA 2003: Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* [PISA 2003: Studies on competence development over a school year]. Münster, Germany: Waxmann.

Prenzel, M., Carstensen, C. H., Schöps, K., & Maurischat, C. (2006). Die Anlage des Längsschnitts bei PISA 2003 [Conceptualization of the PISA 2003 longitudinal study]. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, et al. (Eds.), *PISA 2003: Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (pp. 29–62). Münster, Germany: Waxmann.

Puntambekar, S., & Hübscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist*, *40*(1), 1–12.

Putnam, R. T., Heaton, R. M., Prawat, R. S., & Remillard, J. (1992). Teaching mathematics for understanding: Discussing case studies of four fifth-grade teachers. *The Elementary School Journal*, *93*(2), 213–228.

Reynolds, M. C. (Ed.). (1989). *Knowledge base for the beginning teacher.* New York: Pergamon Press.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55.

Rowan, B., Chiang, F.-S., & Miller, R. J. (1997). Using research on employees' performance to study the effects of teachers on students' achievement. *Sociology of Education*, *70*(4), 256–284.

Scheerens, J., & Bosker, R. (1997). *The foundations of educational effectiveness.* Oxford, UK: Pergamon.

Schilling, S. G. (2007). The role of psychometric modeling in test validation: An application of multidimensional item response theory. *Measurement: Interdisciplinary Research and Perspectives*, *5*(2), 93–106.

Schilling, S. G., & Hill, H. C. (2007). Assessing measures of mathematical knowledge for teaching: A validity argument approach. *Measurement: Interdisciplinary Research and Perspective*, *5*(2), 70–80.

Schmidt, W. H., Tatto, M. T., Bankov, K., Blömeke, S., Cedillo, T., Cogan, L., et al. (2007). *The preparation gap: Teacher education for middle school mathematics in six countries. Mathematics teaching in the 21st century (MT21)*. East Lansing: Michigan State University, Center for Research in Mathematics and Science Education.

Schoenfeld, A. H. (1998). Toward a theory of teaching-in-context. *Issues in Education*, *4*(1), 1–94.

Schoenfeld, A. H. (2007). The complexities of assessing teacher knowledge. *Measurement: Interdisciplinary Research and Perspectives*, *5*(2), 198–204.

Schoenfeld, A. H., Minstrell, J., & van Zee, E. (2000). The detailed analysis of an established teacher carrying out a non-traditional lesson. *Journal of Mathematical Behavior*, *18*(3), 281–325.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, *77*(4), 454–499.

Sfard, A. (2003). Balancing the unbalanceable: The NCTM standards in the light of theories of learning mathematics. In J. Kilpatrick, G. Martin & D. Schifter (Eds.), *A research companion for NCTM standards* (pp. 353–392). Reston, VA: National Council for Teachers of Mathematics.

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer* (*Vol. 1*). Newbury Park, CA: Sage.

Sherin, M. G. (1996). *The nature and dynamics of teachers' content knowledge*. Unpublished doctoral dissertation, University of California, Berkeley.

Shuell, T. (1996). Teaching and learning in the classroom context. In C. D. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 726–764). New York: Simon & Schuster Macmillan.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*(1), 1–22.

Shulman, L. S., & Quinlan, K. M. (1996). The comparative psychology of school subjects. In D. C. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 399–422). New York: Simon & Schuster Macmillan.

Simon, M. A. (1993). Prospective elementary teachers' knowledge of division. *Journal of Research in Mathematics Education*, *24*(3), 233–254.

Smith, E., & Gorard, S. (2007). Improving teacher quality: Lessons from America's No Child Left Behind. *Cambridge Journal of Education*, *37*(2), 191–206.

Stefanou, C. R., Perencevich, K., DiCintio, M., & Turner, J. C. (2004). Supporting autonomy in the classroom: Ways teachers encourage student decision making and ownership. *Educational Psychologist*, *39*(2), 97–110.

Stein, M. K., Baxter, J. A., & Leinhardt, G. (1990). Subject-matter knowledge and elementary instruction: A case from functions and graphing. *American Educational Research Journal*, *27*(4), 639–663.

Stengel, B. S. (1997). "Academic discipline" and "school subject": Contestable curricular concepts. *Journal of Curriculum Studies*, *29*(5), 585–602.

Stigler, J., & Hiebert, J. (2004). Improving mathematics teaching. *Educational Leadership*, *61*(5), 12–17.

Stodolsky, S. S., & Grossman, P. L. (1995). The impact of subject matter on curricular activity: An analysis of five academic subjects. *American Educational Research Journal*, *32*(2), 227–249.

Swafford, J. O., Jones, G. A., & Thornton, C. A. (1997). Increased knowledge in geometry and instructional practice. *Journal for Research in Mathematics Education*, *28*(4), 476–483.

Thompson, A. G., & Thompson, P. W. (1996). Talking about rates conceptually, part II: Mathematical knowledge for teaching. *Journal for Research in Mathematics Education*, *27*(1), 2–24.

Thompson, P. W., & Thompson, A. G. (1994). Talking about rates conceptually, part I: A teacher's struggle. *Journal for Research in Mathematics Education*, *25*(3), 279–303.

Turner, J. C., Meyer, D. K., Cox, K. E., Logan, C., DiCintio, M., & Thomas, C. T. (1998). Creating contexts for involvement in mathematics. *Journal of Educational Psychology*, *90*(4), 730–745.

Tymms, P. (2004). Effect sizes in multilevel models. In I. Schagen & K. Elliot (Eds.), *But what does it mean? The use of effect sizes in educational research* (pp. 55–66). London: National Foundation for Educational Research.

U.S. Department of Education, Office of Policy Planning and Innovation. (2003). *Meeting the highly qualified teachers challenge: The Secretary's second annual report on teacher quality*. Washington, DC: Author.

U.S. Department of Education, Office of Postsecondary Education. (2006). *The secretary's fifth annual report on teacher quality: A highly qualified teacher in every classroom*. Washington, DC: Author.

Vosniadou, S., & Vamvakoussi, X. (2005). Examining mathematics learning from a conceptual change point of view: Implications for the design of learning environments. In L. Verschaffel, F. Dochy, M. Boekaerts & S. Vosniadou (Eds.), *Instructional psychology: Past, present and future trends. Fifteen essays in honour of Erik De Corte* (pp. 55–70). Oxford, UK: Elsevier.

Vosniadou, S., & Verschaffel, L. (2004). Extending the conceptual change approach to mathematics learning and teaching. *Special Issue of Learning and Instruction*, *14*(5), 445–451.

Walberg, H. J., & Paik, S. J. (2000). *Effective educational practices*. Brussels, Belgium: International Academy of Education.

Walshaw, M., & Anthony, G. (2008). The teacher's role in classroom discourse: A review of recent research into mathematics classrooms. *Review of Educational Research*, *78*(3), 516–551.

Wang, M., Haertel, G., & Walberg, H. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, *63*(3), 249–294.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, *54*(3), 427–450.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, *73*(1), 89–122.

Wilson, S. M., & Floden, R. (2003). *Creating effective teachers: Concise answers for hard questions*. Washington, DC: American Association of Colleges for Teacher Education.

Winship, C., & Morgan, S. L. (2007). *Counterfactuals and causal inference: Methods and principles for social research*. Cambridge, UK: Cambridge University Press.

Wu, M. L., Adams, R. J., & Wilson, M. (1997). *ConQuest: Generalized item response modelling software manual*. Melbourne, Australia: Australian Council for Educational Research.

Zeichner, K. M., & Conklin, H. (2005). Research on teacher education programs. In M. Cochran-Smith & K. Zeichner (Eds.), *Studying teacher education: The report of the AERA Panel on Research and Teacher Education* (pp. 645–736). Mahwah, NJ: Erlbaum.

Zumwalt, K. M., & Craig, E. (2005). Teachers' characteristics: Research on the indicators of quality. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying teacher education: The report of the AERA Panel on Research and Teacher Education* (pp. 157–260). Mahwah, NJ: Erlbaum.

Manuscript received June 30, 2008
Final revision received May 28, 2009
Accepted June 13, 2009