

Secondary mathematics teachers' pedagogical content knowledge and content knowledge: validation of the COACTIV constructs

Stefan Krauss · Jürgen Baumert · Werner Blum

Accepted: 5 September 2008 / Published online: 8 October 2008
© The Author(s) 2008. This article is published with open access at Springerlink.com

Abstract Research interest in the professional knowledge of mathematics teachers has grown considerably in recent years. In the COACTIV project, tests of secondary mathematics teachers' pedagogical content knowledge (PCK) and content knowledge (CK) were developed and implemented in a sample of teachers whose classes participated in the PISA 2003/04 longitudinal assessment in Germany. The present article investigates the validity of the COACTIV constructs of PCK and CK. To this end, the COACTIV tests of PCK and CK were administered to various "contrast populations," namely, candidate mathematics teachers, mathematics students, teachers of biology and chemistry, and advanced school students. The hypotheses for each population's performance in the PCK and CK tests were formulated and empirically tested. In addition, the article compares the COACTIV approach with related conceptualizations and findings of two other research groups.

Keywords Pedagogical content knowledge · Content knowledge · Secondary mathematics teachers · COACTIV · Construct validation

1 Introduction

There is broad consensus that teachers' domain-specific knowledge is an essential ingredient of high-quality instruction, particularly in the mathematics classroom (e.g., Ball, Lubienski, and Mewborn, 2001). However, most research on the link between teacher knowledge and instructional outcomes has been either theoretical (e.g., Shulman, 1986, 1987) or based on rather distal indicators of teacher knowledge, such as university grades, number of subject matter courses taken at university, or questionnaire data on beliefs or subjective theories (cf. Hill, Rowan, and Ball, 2005; Pajares, 1992). As a consequence, many questions on mathematics teachers' knowledge, its content, structure, and how it influences teaching and learning, remain open.

Although numerous calls have been made in the literature for valid and reliable assessments of teacher knowledge (e.g., Barnes, 1985; Lanahan, Scotchmer, & McLaughlin, 2004), it was only at the beginning of the twenty-first century that direct tests of mathematics teachers' knowledge were constructed independently by several research groups. In this paper, we report on the professional knowledge tests for secondary mathematics teachers that have been developed in the framework of the COACTIV study. Beside COACTIV, tests on the professional knowledge of mathematics teachers have been developed by Deborah Ball and colleagues in Michigan (targeting elementary teachers in the United States; for results, see, e.g., Hill, Schilling, & Ball, 2004), by the Educational Testing Service (the Praxis series testing candidate mathematics teachers in the United States; see Educational Testing Service, 2006), and within the MT21 project (investigating candidate mathematics teachers and trainee teachers in Germany; see Blömeke, Kaiser, &

S. Krauss (✉) · W. Blum
Department of Mathematics, University of Kassel,
Kassel, Germany
e-mail: skrauss@mathematik.uni-kassel.de

J. Baumert
Center of Educational Research, Max Planck Institute for Human
Development, Berlin, Germany

Lehmann, 2008). All of these approaches can essentially be embedded within Shulman's (1986, 1987) taxonomy of teacher knowledge. Shulman distinguishes theoretically between *pedagogical content knowledge* (PCK), which is the knowledge of "how to make the subject comprehensible to others," and *content knowledge* (CK), which is the "deep understanding of the domain itself." He further identifies *pedagogical knowledge* (PK), which is subject-independent knowledge of how to optimize learning situations in the classroom in general.¹ This last component of teacher knowledge is not addressed in this article.

Construct validity (i.e., the extent to which an operationalization measures the concept it purports to measure) is a crucial issue in psychometric research (Messick, 1988). Other criteria indicating psychometric test quality (i.e., reliability or objectivity) cannot inform on the *meaning* of the constructs measured; in the worst case scenario, interpretations of and conclusions drawn from test results may therefore be invalid. Evidence is thus needed to confirm that the tests of PCK and CK applied indeed measure "pedagogical content knowledge" and "content knowledge" (and not, for example, pedagogical knowledge or general intelligence). Establishing validity means "collecting evidence" (for a critical discussion of the concept, see, e.g., Borsboom, Mellenbergh, & van Heerden, 2004); it differs in this respect from reliability, which can often be calculated and expressed simply by Cronbach's α coefficient (e.g., Nunally & Bernstein, 1994). There are several approaches to validity. For example, the idea of "face validity" in the present context means that the teachers tested should feel that the test items indeed draw on relevant professional knowledge that can be classified as pedagogical content knowledge and content knowledge (subjective validity criterion). Findings showing that the empirical data obtained for PCK and CK are within expectations also testify to the validity of the constructs. For instance, PCK would be expected to predict lesson quality and student learning (i.e., validity in terms of empirical relationships with external criteria).

For the Michigan group, the validity concept is of such importance that a whole issue of *Measurement* (Vol. 5, No. 2–3, 2007) was devoted to examining the validity of the Michigan tests on the professional knowledge of elementary teachers in mathematics (with prominent discussants such as Alan Schoenfeld). Schilling and Hill (2007) suggested an argument-based approach to validity, meaning that the stated assumptions, on the one hand, and their

evaluation in the light of the empirical evidence, on the other, should be strictly separated. The authors note that "despite its importance, test validation is almost universally viewed as the most unsatisfactory aspect of test development," especially because there is a consistent disjunction between theoretical validity conceptualization and validation practice (see also Messick, 1988).

This article investigates the validity of the COACTIV tests of PCK and CK in three steps. In the first section, after briefly introducing the COACTIV project (for an overview, see Kunter et al., 2007), we review empirical findings for both tests (for details, see Brunner et al., 2006a; Krauss et al., 2008a; Krauss et al., 2008b) and discuss the extent to which these results support the validity of our measures of PCK and CK. The validity evidence presented in this first section is primarily in the form of "relationships with external criteria." All analyses presented in this section are based on data obtained *within* the COACTIV study.

In the second section, the main part of the article, we address the issue of construct validity by investigating samples drawn *beyond* the COACTIV study. "Contrast populations" of non-mathematics teachers (candidate mathematics teachers, mathematics students, teachers of biology and chemistry, and advanced school students) were administered the COACTIV tests in an additional study. The basic idea was that if the tests indeed measure what they are supposed to (namely, mathematical PCK and CK), these contrast populations can be expected to show specific patterns of results. Science teachers should score rather low on both tests (especially CK); mathematics students should score high on CK but substantially lower on PCK; candidate mathematics teachers should score higher than the advanced school students but lower than the COACTIV mathematics teachers on both knowledge tests, and so on.

Finally, we compare and contrast the results of this construct validation study with corresponding findings from other research groups. For example, the Michigan group (within the framework of its argument-based approach to validity) has administered its instruments to non-mathematicians and non-teachers; the MT21 project group has investigated the PCK and CK of candidate and trainee teachers.

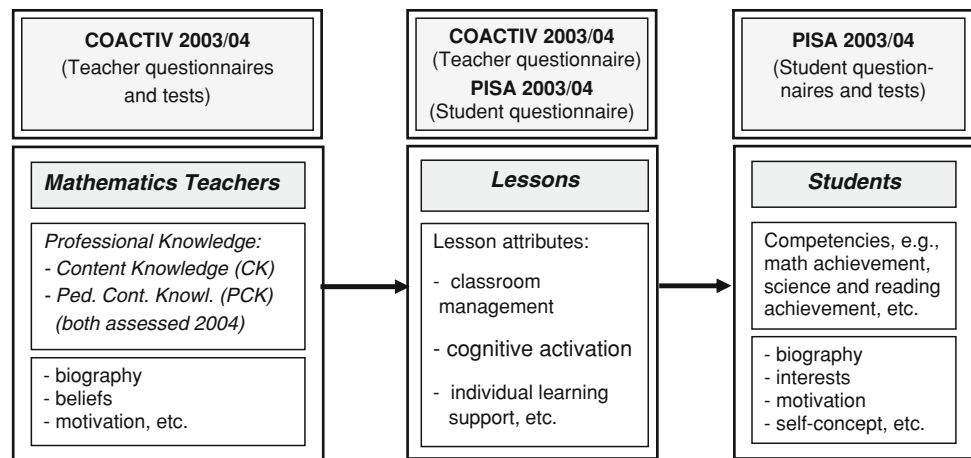
2 The COACTIV tests of PCK and CK

2.1 The COACTIV study

The COACTIV project on *Professional Competence of Teachers, Cognitively Activating Instruction, and the Development of Students' Mathematical Literacy* aimed at conceptualizing and assessing a broad spectrum of teacher

¹ CK, PCK, and PK are today considered the core categories of teacher knowledge (Baumert & Kunter, 2006; Lipowsky, 2006); for additional categories see, e.g., Shulman (1987) or Brunner et al., (2006a).

Fig. 1 Conceptual connection of the COACTIV 2003/04 study and the PISA 2003/04 study with example constructs



competencies, personality variables, and work-related variables in the context of secondary mathematics instruction. The project was funded by the German research foundation (DFG) from 2002 to 2006² (directors: Jürgen Baumert, Berlin; Werner Blum, Kassel; Michael Neubrand, Oldenburg) and surveyed the mathematics teachers whose classes participated in the PISA 2003/2004 longitudinal assessment in Germany (see Prenzel et al., 2004, for details of PISA 2003 and its German extension, and Prenzel et al., 2006, for details of the longitudinal German component).

The close relationship between COACTIV and PISA allows, for the first time in Germany, a combined analysis of large-scale data on teachers, their lessons, and their students within a common technical and conceptual framework (Fig. 1). Whereas the achievements of students and personality variables were assessed in PISA (right column), their teachers were surveyed in COACTIV (left column). Parallel questionnaires on lessons (middle column) were administered to both the students (in PISA) and the teachers (in COACTIV) (“multi-perspectivity”). Note that Fig. 1 depicts only a fraction of the constructs assessed.

On average, the COACTIV 2003/04 teacher assessment took a total of about 12 h, distributed over the course of a school year. Besides knowledge tests, a broad battery of newly developed (or adapted) instruments tapped teachers’ biographical variables, motivational orientations, professional beliefs, and self-regulation (for an overview of the COACTIV instruments, see, e.g., Krauss et al., 2004; Kunter et al., 2007). The students (PISA classes) were administered tests and questionnaires on two school mornings (approx 4 h each). The structure of the data

allows us to use structural equation modeling to test various causal hypotheses, based on the assumption that the teacher influences the lessons, which in turn influence student learning (as indicated by the arrows in Fig. 1). For a general overview of the COACTIV findings, see Kunter et al., (2007) or Brunner et al., (2006b). Detailed results on lessons in the PISA classes from the student and the teacher perspective (Fig. 1, middle column) are reported in Baumert et al., (2004) and Kunter et al., (2005, 2006). Further results from the COACTIV study are reported in Klusmann et al., (2008, on stress and burn out), Kunter et al., (2008, on teacher enthusiasm), Dubberke et al., (2008, on beliefs), Jordan et al., (2008, on the mathematical tasks used in lessons), and Krauss and Brunner (2008, on the competence to react quickly to student answers). In the following, we introduce the core instruments of COACTIV, namely, the tests of secondary mathematics teachers’ PCK and CK. Further details of these tests are given in Krauss et al., (2008b).

2.2 PCK and CK: conceptualization and test construction

2.2.1 Pedagogical content knowledge (PCK)

Shulman (1986) characterizes pedagogical content knowledge as the knowledge needed “to make content comprehensible to others” (p. 9). Taking this as the underlying definition of PCK, we identified three subdimensions that are specifically important to mathematics teaching and used these subdimensions to guide test construction (for details of the theoretical background and the test construction procedure, see Krauss et al., 2008b).

- (1) *Tasks* play a central role in teaching mathematics; much of the time allocated to mathematics lessons is devoted to tasks and their solution. When appropriately selected and implemented, mathematical tasks

² The present empirical validation of the COACTIV constructs of PCK and CK by reference to “contrast populations” is part of a DFG-funded project (2007–2009) conducted by the first author.

lay the foundations for students' construction of knowledge and represent powerful learning opportunities (e.g., Jordan et al., 2008). Because this potential can be exploited by having students consider multiple solutions to specific problems (e.g., Silver, Ghouseini, Gosen, Charalambous, & Font Strawhun, 2005), we assessed teachers' knowledge of tasks by testing their ability to produce multiple solutions. To this end, four items in our PCK test required teachers to list as many different ways as possible for solving a given task.

- (2) Teachers need to work with students' existing conceptions and prior knowledge. Because mistakes can provide valuable insights into the implicit knowledge of the problem solver (Matz, 1982), it is important for teachers to be aware of typical student *misconceptions* and *difficulties*. In our PCK test, this aspect was assessed by presenting teachers with seven scenarios and asking them to detect, analyze (e.g., give cognitive reasons for a given problem), or predict a typical student error or comprehension difficulty.
- (3) Students' construction of knowledge is often only successful with instructional support and guidance; for example, in the form of *explanations* or *representations*. In our PCK test, knowledge of subject-specific instructional strategies was assessed by 11 items that required teachers to explain mathematical situations or to provide useful representations, analogies, illustrations, or examples to make mathematical content accessible to students (see Kirsch, 2000).

Thus, our PCK test contained three subscales: knowledge of mathematical tasks (*Tasks*: 4 items), knowledge of student misconceptions and difficulties (*Students*: 7 items), and knowledge of mathematics-specific instructional strategies (*Instruction*: 11 items). One sample item from each PCK subscale is provided in the Appendix (for more examples of items, see Krauss et al., 2008b).

2.2.2 Content knowledge (CK)

Content knowledge describes a teacher's understanding of the structures of his or her subject. According to Shulman (1986), "the teacher need not only understand that something is so, the teacher must further understand why it is so" (p. 9). Clearly, teachers' knowledge of the mathematical content covered in the school curriculum should be much deeper than that of their students. We conceptualized CK as a deep understanding of the contents of the secondary school mathematics curriculum. It resembles the idea of "elementary mathematics from a higher viewpoint" (in the sense of Klein, 1933). Thirteen items were

constructed to tap teachers' CK in relevant content areas (e.g., arithmetic, algebra, and geometry; see the Appendix for a sample item). No subfacets of CK were assumed (see Krauss et al., 2008a).

Note that this conceptualization clearly distinguishes CK from other possible notions of "content knowledge": (1) the everyday mathematical knowledge that all adults should have, (2) the school-level mathematical knowledge that good school students have, and (3) the university-level mathematical knowledge that does not overlap with the content of the school curriculum (e.g., Galois theory or functional analysis). CK as conceptualized in COACTIV lies between (2) and (3). Because it refers to school mathematics, very good school students might also be expected to solve at least some items.

2.2.3 Other research groups' conceptualizations of PCK and CK

Is the COACTIV conceptualization of PCK and CK coherent and conclusive or might it feasibly be replaced by an entirely different approach? In this section, we compare the COACTIV approach, especially our PCK conceptualization, with the theoretical approaches of other research groups that have recently developed similar assessment instruments. Direct tests of the professional knowledge of mathematics teachers have been developed by the Michigan group (aimed at elementary mathematics teachers) and by the MT21 project group (aimed at teacher students and trainee teachers in middle schools). Because the COACTIV, MT21, and Michigan groups worked independently of each other, substantial overlap between the groups' approaches would testify to their mutual validity at the theoretical level. If a substantial match is found on the conceptual level, the respective empirical results can then be compared (see Sect. 3.5).³

2.2.3.1 The Michigan group

Deborah Ball and colleagues began to discuss ideas on assessing the professional knowledge needed by U.S. elementary teachers in mathematics back in the 1990s within the framework of the Teacher Education and Learning to Teach (TELT) study (e.g., Kennedy, Ball, & McDiarmid, 1993). These efforts are reflected in various theoretical articles focusing on teachers' knowledge (e.g., Ball et al., 2001). Note that the Michigan group uses a somewhat different terminology (basically, the two knowledge categories CK and PCK are subsumed under *mathematical content knowledge needed*

³ The Praxis series (Educational Testing Service, 2006) is not included here because these tests are not used for research purposes (e.g., to inform on populations) but are applied on the individual level (e.g., to certify trainee teachers).

for teaching (MKT), which is distinguished from *subject matter knowledge* or “pure” content knowledge; see also Leinhardt & Greeno, 1986; Sherin, 1996). The following quotation, however, illustrates the conceptual similarities with COACTIV's PCK concept. In elaborating on MKT, Schilling and Hill (2007) specify that “[t]eachers not only need to perform basic computation for themselves, but also need to provide students with explanations for why particular procedures work, to diagnose student errors on those procedures, and to understand non-standard yet correct procedures” (p. 76). Thus, the members of the Michigan group evidently assume teachers to require not only content knowledge, but also knowledge of explanations and knowledge of student errors.

The Michigan group used a matrix of three content areas by three knowledge dimensions as a theoretical framework for developing test items (Hill et al., 2004; Ball, Hill, & Bass, 2005): The three content areas identified were (a) *numbers/operations*, (b) *patterns/functions*, and (c) *algebra*. The three knowledge dimensions were (a) *common knowledge of content* (CKC), which is the mathematical everyday knowledge that all educated adults should have, (b) *specialized knowledge of content* (SKC), which is thought to be teacher specific and acquired only through professional training and classroom experience, and (c) *knowledge of students and content*, a dimension that links mathematical content and student thinking, thus covering knowledge on typical errors or student strategies. This latter dimension thus comes close to COACTIV's PCK.

However, an exploratory factor analysis with a large sample of teachers could not replicate this complex theoretical structure. Instead, it revealed a three-factor solution comprising two content factors (Hill et al., 2004; Schilling & Hill, 2007)—one covering knowledge of patterns, functions, and algebra and another covering knowledge of number concepts and operations—and the PCK factor *knowledge of students and content*. It is interesting that two categories of elementary teachers' content knowledge were distinguished; the same has not been found for secondary teachers (Krauss et al., 2008a; but see Blömeke, Seeber et al., 2008). In a further analysis, Hill et al., (2004) addressed the separation of common content knowledge (CKC) and specialized content knowledge (SKC) by testing a model where, in addition to the three factors explicated, each item was allowed to load on a general factor that was interpreted as the CKC that every adult should have. In this way, they were able to separate CKC from the SKC needed for elementary teaching (which was represented by the three factors identified in the exploratory factor analysis), although the separation was not very distinct. The authors tentatively concluded that their analyses showed evidence of multidimensionality (as opposed

to a single general factor, such as mathematical ability or pure teaching ability), but that a general factor (i.e., common content knowledge) nevertheless operates (for similar analyses on elementary teachers' knowledge of reading, see Phelps & Schilling, 2004). Based on these analyses, the authors developed an IRT-scaled test to assess elementary school teachers' mathematical knowledge for teaching (MKT) that included both common knowledge items and specialized knowledge items. Recently, Hill (2007) has developed an analogous test of middle school mathematics teachers' MKT, with some overlap between the items of the two tests.

In COACTIV, we were also able to distinguish PCK and CK; at the same time, we found evidence that these constructs are closely connected (see Sect. 2.4.2). We return to the Michigan group's tests in Sect. 3.5.

2.2.3.2 The MT21 study The International Association for the Evaluation of Educational Achievement (IEA) is currently conducting an international comparison of the efficiency of teacher education: Learning to Teach Mathematics—Teacher Education and Development Study (TEDS-M). To pilot the study instruments, a pre-study entitled **MT21** (Mathematics Teaching in the 21st Century) was run in eight countries (including the United States and Germany) from 2003 to 2006. MT21 and TEDS-M focus on prospective secondary mathematics teachers. Because MT21 was run in Germany (Blömeke, Kaiser, & Lehmann, 2008) its results are especially suitable for a comparison with COACTIV.

MT21 made a clear theoretical distinction between PCK and CK. The pre-study items of both the PCK and the CK test can be split into subdimensions in two different ways. First, five content areas can be identified, namely, arithmetic, algebra, functions, geometry, and stochastics. Second, the items can be categorized according to the mathematical activities involved, namely, “algorithmatizing,” “problem solving,” and “modeling.” Neither the COACTIV group nor the Michigan group was able to verify this variety of dimensions by factor analytic methods.

It is interesting to note that, as in the items developed by the COACTIV and the Michigan groups, several of the PCK items implemented in MT21 are formulated as “scenarios”: participants are presented with a typical teaching situation and asked to suggest a “didactical solution.” More importantly, inspection of the descriptions of the items administered in the MT21 study clearly reveals that “illustrating” (“Veranschaulichung”) and “students' misconceptions” (“Fehlvorstellungen”) (see Blömeke, Seeber et al., 2008, p. 58f) also play a major role in the theoretical conceptualization of the MT21 approach to PCK.

The approach taken to CK in the MT21 study is also similar to that taken in COACTIV, with distinctions being drawn between school-level mathematical knowledge, school mathematics from a higher viewpoint, and university-level mathematical knowledge (Blömeke, Lehmann et al., 2008, p. 106). However, in the main study (TEDS-M) the trend seems to be to test the mathematical contents that are typically taught in school, but not a deep background understanding of these contents (Tatto et al., 2008). The TEDS-M items distinguish three levels of curricular content knowledge: “novice” (mathematics content that is typically taught in the grades the future teacher will teach), “intermediate” (content that is typically taught one or two grades beyond the highest grade the future teacher will teach), and “advanced” (content that is typically taught three or more years beyond the highest grade the future teacher will teach). Therefore, it should be noted that although the CK tested in TEDS-M will not exceed the level of advanced school knowledge (Schmidt et al., 2007), the MT21 pilot study and COACTIV rely on very similar conceptualizations of CK. We return to the MT21 tests in Sect. 3.5.

2.2.3.3 Summary: the three research groups’ conceptualizations of PCK and CK There is substantial and non-trivial conceptual overlap between the three groups’ approaches (COACTIV, MT21, Michigan), especially with respect to PCK: In accordance with Shulman’s (1986) theoretical characterization, all three groups seem to accept knowledge of explanations and of students’ thinking as the core of mathematics teachers’ pedagogical content knowledge.

Concerning CK, it is difficult to compare the Michigan group’s approach to that of either COACTIV or MT21, because Ball and colleagues are specifically interested in elementary teachers. In particular, their distinction between common content knowledge (CKC) and specialized content knowledge (SKC) is conceptually less useful for the other groups; non-teachers will not be able to solve the items of either the COACTIV CK test (see also Sect. 3) or the MT21 test.⁴

One important formal difference between the approaches must be noted: whereas most of the MT21 and Michigan group items have a multiple choice format (a point critically discussed within both projects, see Blömeke, Kaiser & Lehmann, 2008; Schilling & Hill, 2007),

⁴ Nonetheless, Hill (2007) preserved the distinction between CKC and SKC in her recently developed test of middle school teachers in the United States. Note, however, that U.S. middle schools generally end at grade 8, whereas students in Germany may graduate from secondary school after completing as few as 9 or as many as 13 grades (for details see Sect 1.3).

all PCK and CK items in the COACTIV study have an open-ended format, thus avoiding the problems typically associated with multiple choice items (e.g., guessing; Millman, Bishop, & Ebel, 1965).

2.3 Test implementation in COACTIV: sample and procedure

The teachers participating in COACTIV 2004 taught mathematics in the 10th grade classes sampled within the framework of PISA 2003/2004 in Germany. Our teacher sample can thus be considered fairly representative of German 10th grade mathematics teachers. The COACTIV instruments were administered at two measurement points corresponding to the dates of the German PISA assessments in April 2003 (9th grade) and April 2004 (same classes; 10th grade). A total of 218 secondary mathematics teachers participated at the second COACTIV measurement point (2004), when the tests of PCK and CK were implemented; 198 teachers completed both tests.

For several of the subsequent analyses, these 198 teachers were split into two groups. The rationale for this distinction lies in the structure of the German secondary school system. Students in Germany are tracked to different secondary school types at the age of 10 to 11 years (end of 4th grade), based largely on their educational attainment to date. The 16 federal states implement between two and four secondary tracks, the most academic being the *Gymnasium*. The major difference between the tracks is that *Gymnasium* students are college bound, whereas the other tracks are more vocationally oriented.⁵ Teacher candidates in Germany must have graduated from *Gymnasium*, regardless of the school type in which they aspire to teach. However, teacher candidates training for the academic track complete a 4- to 5-year phase of university-based training (first phase of teacher education) plus a 2-year compulsory teaching placement in a school (second phase of teacher education), whereas those training for the non-academic tracks study for 3 to 4 years at teacher college or university, followed by a 2-year compulsory teaching placement. The practice-oriented compulsory teaching placement, during which teacher candidates are responsible for their own classes for the first time, is comparable between tracks, but the university phase differs substantially: Teacher candidates aspiring to teach mathematics in the academic track study the subject at a much deeper and more theoretical level, to some extent comparable to students

⁵ Depending on the federal state, the academic track *Gymnasium* ends at grade 12 or 13; the vocationally oriented tracks end at grade 9 or 10.

Table 1 CK and PCK: means M (standard deviations SD) and empirical maxima by teacher group

	M (SD) GY ($N = 85$)	M (SD) NGY ($N = 113$)	Effect size d (GY vs. NGY)	Emp. max. GY	Emp. max. NGY
CK (13 items)	8.5 (2.3)	4.0 (2.8)	1.73	13	12
PCK (22 items)	22.6 (5.9)	18.0 (5.6)	0.80	37	29
Instruction (11 items)	9.3 (3.4)	7.1 (3.2)	0.67	17	15
Students (7 items)	5.8 (2.3)	4.3 (1.9)	0.71	11	9
Tasks (4 items)	7.5 (1.8)	6.6 (2.0)	0.47	12	10

GY academic track teachers, NGY non-academic track teachers. According to Cohen (1992), $d = 0.20$ is a small effect, $d = 0.50$ a medium effect, and $d = 0.80$ a large effect. All differences are significant at $p < 0.01$

majoring in mathematics; those training for the other secondary tracks receive more varied general and practical pedagogical training. The teachers of the COACTIV sample were therefore divided into two subgroups—teachers in the academic track “GY” versus teachers in the non-academic tracks “NGY”⁶—for some of the statistical analyses (see Krauss et al., 2008b, for information on the distribution of the COACTIV NGY subsample across the non-academic tracks).

Of the 198 teachers, 85 (55% male) taught in the academic track (GY) and 113 (43% male) in other secondary school types (NGY). The average age of participating teachers was 47.2 years ($SD = 8.4$). Teachers were paid 60 euro for participation. The assessment of PCK and CK was conducted individually in a separate room at the teacher's school in the afternoon of the day their PISA students were tested. It was administered as a power test with no time constraints by a trained test administrator. The teachers were not allowed to use a calculator. The average time required to complete the 35 items was about 2 h (approx 65 min for the 22 PCK items and 55 min for the 13 CK items). In terms of face validity, the teachers' evaluation of the relevance of the items was positive (e.g., one teacher wrote: “I know I should know this”).

All 35 items were open-ended. A scoring scheme was developed and eight raters were given extensive training. The responses to each test item were coded by two raters independently. The inter-rater objectivity ρ (Shavelson & Webb, 1991) was very satisfactory (on average across all items, ρ was 0.81). Furthermore, both tests yielded satisfactory reliabilities (Cronbach's alpha was 0.78 for PCK and 0.83 for CK). Thus, in terms of objectivity and reliability, the test construction can be considered successful. In the following, we review the main results and discuss consequences for the validity of the underlying knowledge constructs.

⁶ Because the word “Gymnasium” invites false associations, we avoid it in the following, referring instead to the “academic track.” In the abbreviation “GY” however, the original name is preserved.

2.4 Results

2.4.1 Means and school type differences

The largest source of variance in teachers' performance was whether or not they taught in the academic track. As shown in Table 1, there were very large differences in CK ($d = 1.73$; see caption of Table 1) and large differences in PCK ($d = 0.80$) with respect to school type, both indicating higher expertise among teachers in the academic track (GY).

The large difference in CK reflects the intensive coverage of mathematical subject knowledge in GY teachers' university training. However, their advantage in PCK, especially in the student and instruction subscales, is remarkable, given that GY teachers usually receive *less* training in the teaching of the subject (“Fachdidaktik,” i.e., pedagogical content knowledge) and in pedagogy (or educational psychology) at university. Yet this finding is in line with the results of many qualitative studies (e.g., Baumert & Kunter, 2006) that point to a close relationship between PCK and CK (see also Sect. 2.4.2). Finally, it should be noted that Brunner et al., (2006a) showed that, when CK is statistically controlled (i.e., when only teachers with the same CK level are compared), the NGY teachers slightly outperform the GY teachers in terms of PCK.

2.4.2 Relationship between PCK and CK

The relationship between the two knowledge categories can be examined directly by calculating the manifest bivariate correlation between PCK and CK, which in the COACTIV data was 0.60. Note that this connection was much stronger in the GY group; indeed, modeling PCK and CK as latent constructs led to a latent correlation in the GY group that was no longer statistically distinguishable from 1 (see Krauss et al., 2008a). Despite this high correlation, however, the effect sizes between the two groups of teachers with respect to the two knowledge categories differed markedly ($d = 1.79$ for CK vs. “only” 0.80 for PCK).

Why was this correlation less strong in the NGY group? Closer inspection of the teacher data revealed that some NGY teachers who performed very poorly on CK (e.g., scoring only 1–2 points) nevertheless showed above-average performance on PCK. In other words, although our data support the claim that PCK profits from a solid base of CK, CK is only *one* possible route to PCK. The greater emphasis on didactics in the initial training provided for NGY teacher candidates in Germany may be another route.

2.4.3 Knowledge and working experience

Interestingly, no positive correlations were found between either of the knowledge categories and years of professional experience as a teacher (see Brunner et al., 2006b; Krauss et al., 2008b). These findings indicate that teachers' knowledge no longer seems to develop a great deal (at least in terms of the COACTIV items) once they have completed their training. This finding seems surprising; it contradicts theories that attribute teachers' expertise development explicitly to their practical experience (Hashweh, 2005; Hiebert, Gallimore, & Stigler, 2002). According to deliberate practice theory, however, expertise does not increase simply by doing a job (Ericsson, Krampe, & Tesch-Römer, 1993). Rather, motivation and *deliberate* practice is required to identify and overcome one's weaknesses, preferably with the support of ongoing expert feedback. Because these conditions are normally not given in everyday school life (in contrast to teacher training, see below), our findings are in line with deliberate practice theory, the predictions of which have already been verified for various other domains (e.g., music, sports, medicine, chess, etc.).

2.4.4 Knowledge and subjective beliefs

Kunter et al., (2007) and Dubberke et al., (2008) analyzed the relations of PCK and CK with teachers' subjective beliefs on the nature of mathematics and on the learning of mathematics. They found, for example, that teachers with high PCK and CK scores tended to disagree with the view that mathematics is “just” a toolbox of facts and rules that “simply” have to be recalled and applied. Rather, these teachers tended to think of mathematics as a process permanently leading to new discoveries. At the same time, the knowledgeable teachers rejected a receptive view of learning (“mathematics can best be learned by careful listening”), but tended to think that mathematics should be learned by self-determined, independent activities that foster real insight. These relationships between knowledge and beliefs nicely fit into the desirable “profile” (Sternberg and Horvarth, 1995) of an “expert teacher” (Palmer, Stough, Burdenski, & Gonzales, 2005).

2.4.5 Knowledge and student learning progress

Because COACTIV was “docked” onto the PISA study, it was possible to relate teachers' PCK to their students' mathematics achievement gains over the year under investigation. Very briefly, when their mathematics achievement in grade 9 was kept constant, students taught by teachers with higher PCK scores performed significantly better in mathematics in grade 10. By means of structural equation modeling Baumert et al. (2006, 2008) could show that PCK, mediated by aspects of the lesson, can explain students' achievement gains in a non-trivial way.

Because these relations were much weaker for CK, our results demonstrate that PCK is indeed a necessary prerequisite for teachers being able to create powerful learning environments that support their students' learning. Because student learning can be considered the ultimate aim of teaching, this finding is a strong indicator of the (predictive) validity of PCK as conceptualized and operationalized in COACTIV.

3 Construct validation by reference to contrast populations

In this section, we examine the validity of both knowledge constructs by going beyond the COACTIV data and administering our tests of PCK and CK to theoretically specified contrast populations. The rationale behind this approach was as follows: if the COACTIV tests indeed measure secondary mathematics teachers' pedagogical content knowledge and content knowledge, it should be possible to formulate hypotheses regarding the performance of other populations on these tests. For example, teachers of biology and chemistry can be expected to score rather low on both tests (especially on the CK test), whereas subject matter specialists can be expected to score relatively high on CK, but much lower on PCK. At the same time, knowledge of both areas can be expected to increase continuously during teacher training. Therefore, mathematics teacher candidates can be expected to score higher than (even advanced) school students on both knowledge categories, but lower than the in-service COACTIV teachers. In order to provide a theoretical framework for our investigation of contrast populations, we first introduce two complementary hypotheses, namely the Professional Knowledge Hypothesis and the Growing Knowledge Hypothesis.

3.1 Professional knowledge hypothesis

The highly specialized professional knowledge of teachers is considered to be one of the main features distinguishing them from laypeople (see the German debate on teacher

professionalization, e.g., Bromme, 1992; for professions in general, see Mieg, 2001). The easiest way of testing this professional knowledge hypothesis would be to administer the tests of PCK and CK to a random sample of adults. Because most respondents in such a sample would probably not be able to solve a single item, however, we would not learn much from this approach. Instead we chose a more conservative approach and investigated “related professionals.” Mathematics teachers are professionals on at least two dimensions: they are both professional mathematicians and professional teachers. Our choice of contrast populations for testing the professional knowledge hypothesis was thus informed by varying these two dimensions of professionalism independently (Table 2).

According to Ackerman's (1996) *theory of adult intellectual development*, two types of tests are needed to provide a representation of an adult's knowledge: “a *deep* test of professional knowledge, and a broad array of more *shallow* tests outside the profession” (p. 241). From this viewpoint, the professional knowledge hypothesis aims at analyzing which of the two knowledge categories is *deeply* ingrained in the populations investigated.

In the following, we elaborate on the groups specified in the cells of Table 2 and formulate hypotheses regarding their performance on the COACTIV tests of PCK and CK.

3.1.1 Sample 1: COACTIV teachers

The COACTIV mathematics teachers are introduced in Sect. 2.3. The performance of the COACTIV mathematics teachers (as discussed in Sect. 2.4) can serve as a limit of expectations for the other groups' results.

3.1.2 Sample 2: biology/chemistry teachers (GY)

Physics teaching is clearly the profession most closely related to mathematics teaching. However, it is hard to find teachers of physics who are not at the same time teachers of mathematics. Moreover, the professional knowledge of physics teachers is so strongly rooted in mathematics that they do not qualify as a contrast population. On the other hand, teachers of languages (or music, arts, religion, etc.,) would probably not be able to solve the mathematics items. It therefore seemed reasonable to choose other science teachers, namely, teachers of biology and chemistry, whose

university training covered some aspects of mathematics, but who do not use mathematics in their everyday teaching to the same extent as mathematics or physics teachers. Again taking a conservative approach, we chose teachers in the academic track who had studied and taught both biology *and* chemistry. We hypothesized that these teachers would score low on mathematical PCK and even lower on mathematical CK.

3.1.3 Sample 3: students majoring in mathematics

The obvious idea for this cell of Table 2 (subject matter specialists) would be to investigate professional mathematicians. Because they work in various fields (e.g., industry, research, insurance companies), however, the professional development of their knowledge after university is highly variable. We therefore chose to investigate students majoring in mathematics toward the end of their university career. Not only do these students constitute a more homogeneous group, they are also easier to recruit and to examine in groups. Furthermore, it is possible to analyze the direct impact of their university training on their PCK and CK (without the influence of their subsequent professional experience, which may vary dramatically).

We hypothesized that the CK of mathematics students would be comparable to that of the GY teachers, but that their PCK scores would be considerably lower. Given the particularly strong correlation between the two knowledge categories found for teachers in the academic track (Sect. 2.4.2), however, the mathematics students might alternatively be expected to score high on PCK as well.

3.1.4 Sample 4: GY school students in advanced grade 13 mathematics courses

The final cell in Table 2 could be filled with a random sample of adults. To provide more informative results, we chose 18–19-year-old students enrolled in advanced mathematics courses in grade 13. This kind of pre-university course only exists in the academic track, where students can specialize in certain subjects in the upper secondary years. Of all populations without university training, this group has the highest mathematical expertise. At the same time, the participants still are very close to the field of interest (curriculum-oriented content knowledge

Table 2 Professional knowledge hypothesis: two dimensions of mathematics teachers' professionalism and the corresponding contrast populations (samples 2–4)

	Mathematician	Non-mathematician
Teacher	Sample 1: COACTIV teachers (in-service mathematics teachers)	Sample 2: Biology/chemistry teachers (GY)
Non-teacher	Sample 3: Students majoring in mathematics (end of university education)	Sample 4: GY school students (in advanced grade 13 mathematics courses)

Table 3 Growing knowledge hypothesis: three stages on the path to becoming a mathematics teacher and the corresponding samples

School students		Mathematics teacher candidates		Mathematics teachers
Sample 4	→	Sample 5	→	Sample 1 _{GY}
GY school students (in advanced grade 13 mathematics courses)		GY mathematics teacher candidates (end of university education)		COACTIV GY teachers

and pedagogical content knowledge). Thus, we hypothesized that they would be able to solve some of the items from both the CK and the PCK tests.

3.2 Growing knowledge hypothesis

The growing knowledge hypothesis states that PCK and CK (as opposed to personality traits, such as intelligence) develop continuously during the process of teacher training and professionalization (for the teaching profession in particular, see, e.g., Berliner, 2001, or Sternberg & Horvath, 1995; for general considerations, see Ericsson, Krampe, & Tesch-Römer, 1993, or Mieg, 2001). Samples 4 and 1 mark the starting point and the end point of this process of professionalization (Table 2). Because many (but not all) mathematics teachers were previously enrolled in advanced mathematics courses at upper secondary level, sample 4 can be used to approximate the (maximum possible) starting level of PCK and CK before students enter university. At the other end of the continuum, sample 1 (COACTIV teachers) informs on the PCK and CK of in-service teachers. To complete the design, we examined a connecting link between school students and COACTIV teachers, namely, mathematics teacher candidates at the end of their first phase of teacher education (see Table 3).

Because previous findings have shown that the PCK and CK of the teachers do not improve with years of classroom experience (see Sect. 2.4.3), in the framework of the growing knowledge hypothesis we focus on the pre-service training of mathematics teachers. Both the university training phase and the subsequent 2-year teaching placement at school satisfy the “deliberate practice” conditions (Ericsson et al., 1993) for the development of expertise; regular examinations motivate teacher candidates to improve both their professional knowledge and their teaching expertise as well as to overcome their weaknesses and knowledge gaps, while supervisors and examiners provide regular expert feedback. During both phases of training, the candidates’ profession is *learning* (and not yet *teaching*); teacher education can therefore be considered an ideal platform for *deliberate* practice of both PCK and CK.

It must be acknowledged that cross-sectional data allow only a “dirty” approximation of real growing processes (cf. Keeves, 1992). At present, however, there is a dearth of empirical research on candidates’ knowledge levels at the

different stages of teacher education (but see Blömeke, Kaiser, & Lehmann, 2008). The initial findings on the differences between the three samples will help us to develop appropriate longitudinal designs for future research. Where the middle column of Table 3 is concerned, we have to date tested only mathematics teacher candidates aspiring to teach in the academic track (the other groups are currently under investigation). Therefore, we limit our examination of the growing knowledge hypothesis to the academic track, restricting the sample of COACTIV teachers (right column of Table 3) to teachers in the academic track (sample “1_{GY}”) in these analyses.

The samples of school students in advanced mathematics courses and of COACTIV teachers were introduced in Sect. 3.1. In the following, we describe the remaining sample of mathematics teacher candidates.

3.2.1 Sample 5: mathematics teacher candidates (academic track)

As described above, the subject matter university training provided for students aspiring to teach in the academic track in Germany is comparable to that provided for subject matter students (sample 3: students majoring in mathematics), at least in the first half of their studies.⁷ It is important to note that teacher candidates also have to study a second subject at the same time (mathematics teacher candidates often choose physics). We chose teacher candidates approaching the end of their university education (thus allowing direct comparison with sample 3). Based on the growing knowledge hypothesis, we expected teacher candidates to score lower than the COACTIV teachers in the academic track (sample 1_{GY}) on both knowledge categories, but considerably higher than the school students (sample 4). We further explored whether teachers acquire more of their PCK and CK in the first phase of teacher education at university (in which case the difference between sample 4 and sample 5 would be the larger one) or in the second phase in schools (in which case the difference between sample 5 and sample 1 would be the larger one).

⁷ The 16 German states vary somewhat with respect to this comparability. In all states, however, students aspiring to teach at the academic track have to study the subject matter in considerably more depth than their peers aspiring to teach at the other secondary school types (e.g., *Realschule* or *Hauptschule*).

Furthermore, the teacher candidates were expected to score slightly lower in CK than the students majoring in mathematics, but to outperform them in PCK.

3.3 Samples and procedure

In the following, we briefly describe the samples drawn and outline the procedure of test administration in each group (for sample 1, the COACTIV teacher sample, see Sect. 2.3). It should be emphasized that all samples were recruited by voluntary participation via announcements in the participants' institutions. Consequently, the samples may not be fully representative, and the results must therefore be interpreted as indicative findings that might help to develop more specifically formulated hypotheses rather than as conclusive findings.

3.3.1 Sample 2: biology/chemistry teachers (GY)

Biology and chemistry teachers were extremely hard to convince as to the (scientific) benefits of their completing a test of *mathematical* PCK and CK. They were therefore offered 50 euro (approx. US\$75) for participation (double the compensation offered to the other participants). In total, 16 biology and chemistry teachers from different academic track Berlin schools (all of whom were trained in and taught both biology and chemistry) were administered the COACTIV tests of PCK and CK; 12 (75%) were female and their average age was 49.1 years (SD: 6.9).

3.3.2 Sample 3: students majoring in mathematics

A sample of 137 students majoring in mathematics were recruited from three Berlin universities (Free University, Humboldt University, and Technical University) and from the universities of Potsdam, Dresden, Erlangen-Nuremberg, and Kassel. All students were tested in small groups in their universities and paid 25 euro for participation. Of the participating students, 87 (63.5%) were male and the average age was 23.9 years (SD: 1.9). On average, they had been enrolled at university for 6.4 semesters (SD: 1.9).

3.3.3 Sample 4: GY school students (in advanced grade 13 mathematics courses)

The PCK and CK instruments were administered to 30 students enrolled in advanced mathematics courses in three academic track Berlin schools. They were tested in their schools in groups of 6, 9, and 15 students and paid 25 euro for participation. Of the students, 20 (67%) were male and the average age was 18.6 years (SD: 0.7).

3.3.4 Sample 5: GY mathematics teacher candidates

A sample of 90 teacher candidates aspiring to teach mathematics in the academic track were recruited from three Berlin universities (Free University, Humboldt University, and Technical University), and from the universities of Potsdam, Dresden, and Kassel. They were tested in small groups in their universities and paid 25 euro for participation. Of the teacher candidates, 37 (41%) were male and the average age was 25.2 years (SD: 2.2). On average, they had been enrolled at university for 7.7 semesters (SD: 2.4).

As for the COACTIV teachers (sample 1), the procedure for test administration in samples 2 to 5 was as follows: The tests were administered by a trained test administrator in the participants' institution. There were no time limits, and the participants were not allowed to use a calculator. In addition to the PCK and the CK tests, all participants were administered a questionnaire assessing biographical background variables and their experience of teaching mathematics (e.g., whether and how often they gave extra lessons in mathematics, etc.). The questionnaire administered to samples 2 to 5, however, was much shorter than that administered to the COACTIV teachers.

3.4 Results

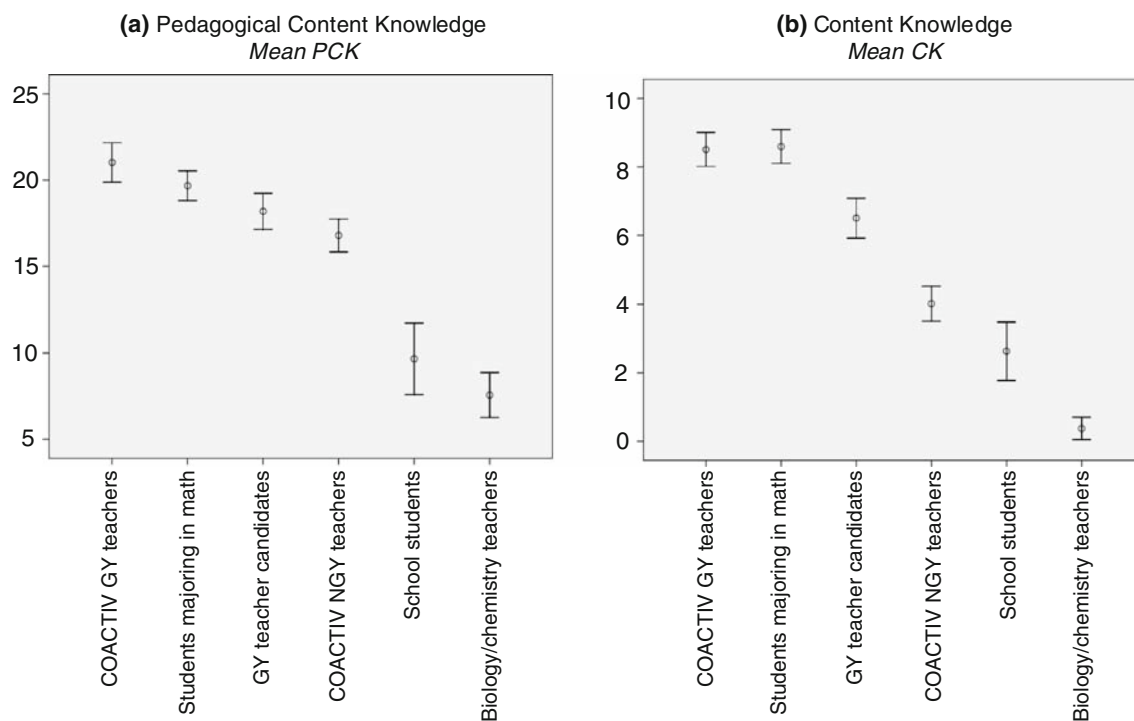
3.4.1 General overview

Before discussing the findings in detail, we first summarize all results for samples 1–5 in Table 4 (for visualization, see Fig. 2a, b).

First note that the results for the COACTIV teachers in Table 4 are slightly different from those presented in Table 1, the reason being that in the COACTIV study two items in the PCK instruction subscale were assessed using computer-based measures (geometrical animations were displayed). These items were not administered to the other samples for logistical reasons; the COACTIV teachers' scores on these items were therefore excluded from the results displayed in Table 4 (this does not, however, substantively influence any of our findings reported above). Across all samples ($N = 471$), Cronbach's α was 0.80 for PCK (20 items) and 0.85 for CK (13 items). It should be noted that the results for samples 2 and 4, in view of their small sample sizes, must be treated with caution. Figure 2a displays the PCK scores and Fig. 2b the CK scores of all samples (rank-ordered according to the PCK score). The error bars represent 95% confidence intervals (according to Cumming and Finch, 2005, two samples differ significantly at, for example, $p < 0.01$ if the corresponding intervals do not overlap).

Table 4 PCK and CK: means M (and standard deviations SD) for all samples

Sample no.	N		Pedagogical content knowledge (PCK) M (SD)	Content knowledge (CK) M (SD)
1	198	COACTIV 2004 teachers	18.6 (5.6)	5.9 (3.4)
I_{GY}	85	COACTIV 2004 teachers (GY)	21.0 (5.3)	8.5 (2.3)
I_{NGY}	113	COACTIV 2004 teachers (NGY)	16.8 (5.1)	4.0 (2.8)
2	16	Biology/chemistry teachers (GY)	7.6 (2.5)	0.4 (0.6)
3	137	Students majoring in mathematics (end of university education)	19.7 (5.1)	8.6 (3.0)
4	30	GY school students (in advanced grade 13 mathematics courses)	9.7 (5.6)	2.6 (2.3)
5	90	GY mathematics teacher candidates (end of university education)	18.2 (5.0)	6.6 (2.8)

**Fig. 2** PCK and CK scores of all samples considered in the construct validation study (rank-ordered according to PCK score). The *error bars* represent 95% confidence intervals

A glance at Table 4 and Fig. 2a and b reveals that all relationships—with the exception of the mathematics students' mean PCK score of 19.7—were in the range of our expectations. The PCK scores of students majoring in mathematics appear so striking, however, that they deserve to be considered separately (see Sect. 3.4.4). Let us first discuss the other results in terms of their support for the two hypotheses formulated.

3.4.2 Professional knowledge hypothesis

The grade 13 students in advanced mathematics courses (sample 4) and the biology/chemistry teachers (sample 2) showed comparable levels of PCK (9.7 and 7.6,

respectively; see Fig. 2a), well below the level reached by the mathematics teachers (sample 1; 18.6). The two samples differed in terms of CK; however, with the school students scoring somewhat higher than the biology/chemistry teachers (2.6 and 0.4, respectively; see Fig. 2b). This result can be attributed to the curriculum-oriented conceptualization of CK in COACTIV. As expected, the mathematics students' (sample 3) performance on the test of CK (8.6) was comparable to that of the GY teachers (8.5). This result is basically in line with Shulman's (1987) assertion: "We expect that the subject matter understanding of the teacher be at least equal to that of his or her lay colleague, the mere subject matter major" (p. 8). A glance at Fig. 2a reveals that samples I_{GY} , I_{NGY} , 3, and 5 do have

Table 5 Test of the growing knowledge hypothesis (academic track)

	Sample 4 GY school students (in advanced grade 13 mathematics courses)	Sample 5 GY teacher candidates (end of university education)	Sample 1 _{GY} COACTIV GY teachers
CK	2.6	6.6	8.5
PCK	9.7	18.2	21.0

“deep” mathematical PCK, whereas samples 2 and 4 do not. Figure 2b shows a similar pattern of results, with the NGY teachers showing a relatively low level of performance (a finding that can be explained by the structure of the German teacher education system, see 2.3). Thus far, the findings are in line with the professional knowledge hypothesis (with the exception of the mathematics students' performance on PCK, which is discussed below).

Note that, interestingly, the ratio of PCK divided by CK is within the relatively small range of 2.5–4.2 for all samples except one: the ratio for the biology/chemistry teachers is about 19, indicating that this group has an extraordinarily high level of mathematical PCK in view of their poor CK. Given that biology/chemistry teachers are the only contrast population with general pedagogical knowledge PK, this finding is congruent with Shulman's “amalgam” hypothesis, which basically states that PK in combination with CK “amalgamates” to form PCK (Shulman, 1987). The same argumentation may apply to some extent to the NGY teachers, who have the second highest ratio of PCK to CK (namely 4.2). As NGY teachers are exposed to less CK and more PK in the university-based phase of their teacher training, it seems that they draw substantially on their PK to develop PCK (but without fully compensating for their lack of CK).

3.4.3 Growing knowledge hypothesis

Table 5 presents the samples examined in our test of the growing knowledge hypothesis.

The highest possible level of mathematical PCK and CK acquired before university entrance can be approximated by the performance of sample 4. Taking sample 4 as the starting point and sample 1 as the end point of the process of teacher professionalization, it is clear from Table 5 that roughly two-thirds of teacher candidates' knowledge gains (in both PCK and CK) can be attributed to their university training (given that not all teacher candidates attended an advanced mathematics course, this effect might in fact be even larger). Because PCK and CK do not show further improvement with years of classroom practice (Sect. 2.4.3), it may be speculated that the remaining third can be attributed to the second phase of pre-service training (i.e., the 2-year compulsory teaching placement).⁸ Data from

trainee teachers in their second phase of teacher education are needed to address this point more specifically. In order to fill this missing link in Table 5, the COACTIV-R longitudinal study is currently assessing trainee teachers at two measurement points in this second phase of teacher education.

Although we do not yet have test data from teacher candidates aspiring to teach at other school types (NGY), we expect their knowledge gains in CK during teacher training to be less pronounced (the data for qualified NGY teachers presented in Table 4 suggest an increase from 2.6 to 4.0). Acknowledging the role of CK as a prerequisite for PCK, their expected gain in PCK (from about 9.7 to 16.8) is therefore very respectable. Comparison of the increase in NGY teachers' PCK and CK (relative to school students; Fig. 2a vs. Fig. 2b), however, reflects the fact that teacher education for NGY teachers in Germany focuses more on PCK than on CK.

An alternative interpretation of the increase depicted in Table 5, especially in CK, is that some of the items solved by the GY teachers were simply not feasible for school students. Interestingly, however, inspection of the data showed that not a single PCK item and only 1 of the 13 CK items was not solved by at least one school student. Furthermore, all PCK and CK items could be solved by several teacher students. In principle, the COACTIV items were thus within the reach of teacher students and even of very good school students.

In sum, the data not only support the hypothesis of continuous improvement of PCK and CK during teacher training and professionalization, they also give some indication of the shape of the curve of knowledge growth.

3.4.4 The unexpectedly high PCK scores of students majoring in mathematics

To understand the unexpectedly high PCK scores of mathematics students, we first split the PCK scores into the subfacets of Tasks, Students, and Instruction and contrasted

⁸ There is some evidence, however, that especially in the first few years as a fully qualified teacher, professional expertise increases substantially. Unfortunately, this question cannot be addressed using the COACTIV data, because we have too few cases at this stage of the teaching career.

Table 6 Performance in CK and in PCK and its subfacets: comparison of COACTIV GY teachers and students majoring in mathematics

	COACTIV GY teachers (sample 1 _{GY}) M (SD)	Students majoring in mathematics (sample 3) M (SD)	Effect size <i>d</i>
Content knowledge (CK)	8.5 (2.3)	8.6 (3.0)	−0.03
Pedagogical content knowledge (PCK)	21.0 (5.3)	19.7 (5.1)	0.26
<i>Tasks</i>	7.5 (2.3)	7.1 (1.6)	0.19
<i>Students</i>	5.8 (1.8)	6.0 (1.8)	−0.08
<i>Instruction</i>	7.8 (2.7)	6.6 (3.0)	0.40*

* $p < 0.01$

the mathematics students' scores with those of the COACTIV GY teachers (Table 6).

We chose to compare PCK in the samples of GY teachers and mathematics students because CK can be assumed to be the same in both groups. As shown in Table 6, the GY teachers' advantage in PCK can be attributed primarily to the "Instruction" subfacet; the teachers in the academic track scored significantly higher than the mathematics students on only the Instruction items, indicating that this subdimension may be a core aspect of pedagogical content knowledge. Indeed, this is the most lesson-related subfacet—knowledge on *content* and knowledge on *students* have to be integrated to produce an interactive teaching decision on how to proceed.

Note, however, that although the GY teachers outperformed the mathematics students in terms of PCK, the mathematics students in turn outperformed both the NGY teachers and the teacher candidates. Does this mean—to put it quite simply—that subject matter students should be recruited for schools, or, alternatively, that teacher training should be aligned to that of subject matter specialists? Or does it mean that the PCK test in fact measures something other than PCK (but what?). Before we answer these questions, the following issues should be noted:

- (1) In contrast to previous pedagogical or psychological approaches (cf. Ball et al., 2001; Mayer, 2004), our test of PCK was by definition subject-oriented. A deep understanding of mathematics *should* support teachers' pedagogical content knowledge substantially (also see Shulman's, 1986, idea of PCK as an amalgam of CK and PK).
- (2) The mathematics students are a particularly selective sample. They were recruited in selected universities and participation was voluntary. It can be assumed that the students who chose to participate expected to be able to solve the "pedagogical content knowledge" items they had been told would be administered.
- (3) Students majoring in mathematics generally have slightly higher cognitive abilities (IQ) than teacher candidates. However, statistically adjusting for this difference (which can be roughly approximated by

the participants' *Abitur* grades; see Baron-Boldt, Schuler, & Funke, 1988) did not make a substantial difference to the results (the average *Abitur* grade [GPA] of mathematics students was 1.8; that of the teacher candidates was 2.0; *Abitur* grades are calculated on a scale of 1.0–6.0, with 1.0 being a perfect score and 4.0 being the pass mark).

- (4) Teachers and teacher candidates outperformed mathematics students in the geometry items. Geometry plays a major role in schools, but only a minor role at university. Although the COACTIV tests are curriculum oriented, they are dominated by algebra and arithmetic; geometry is rather underrepresented (an issue that will be addressed in future test development). Overall, the GY teachers outperformed mathematics students on the 8 geometry items (both PCK and CK) with an effect size of $d = 0.32$.
- (5) Our laboratory test does not indicate whether participants are actually able to capitalize on their knowledge in real lessons. The PCK test can only measure the theoretical competence that participants *might* exploit in lessons.
- (6) Most importantly, it must be acknowledged that teachers (and teacher candidates) have all this knowledge *twice*: in addition to mathematics, they study and teach a second subject at the same time. From this perspective, the teacher candidates' performance relative to that of the subject matter specialists is quite remarkable. Indeed it seems reasonable to ask why students majoring in mathematics, who devote nearly their entire study time to the subject of mathematics, do not outperform teachers and teacher candidates (who devote only half of their time to mathematics) more clearly.

All these aspects suggest that it would not be justified at all to simply hire subject matter specialists for schools in the expectation that their content knowledge would automatically enable them to deliver high-quality teaching that would in turn foster student learning. Rather, the data support the hypothesis that it is a *combination* of content knowledge, general pedagogical knowledge, and the ability

to actually apply this knowledge in the classroom that accounts for teachers' effectiveness.

3.5 Corresponding findings of the Michigan group and the MT21 project group

In Sect. 2.2.3, we introduced the tests on the professional knowledge of mathematics teachers developed by the Michigan group and the MT21 project group, comparing them with the COACTIV tests at a theoretical level. In the following, we consider the empirical findings of the Michigan group and the MT21 project group, highlighting commonalities and differences with our construct validation study (Sect. 3.4).

3.5.1 The Michigan group

Similarly to the COACTIV group (Baumert et al., 2008), Hill et al., (2005) have verified the effects of teacher knowledge on student learning, thus providing strong validation of their instruments. Given that the two samples investigated (elementary teachers vs. secondary mathematics teachers) differed considerably in terms of their mathematical expertise, this demonstrates that the overlapping conceptual paths taken by the Michigan group (MTK) and by COACTIV (PCK) seem to tap the core business of teaching, regardless of the grade specified.

A whole issue of *Measurement* (Vol. 5, No. 2–3, 2007) further addresses the validity of the Michigan group's tests. Hill, Dean, and Goffney (2007) presented a selection of their items to non-teachers and to mathematicians. However, they were not so much interested in the percentages of correct answers to their multiple choice items as in the participants' way of thinking. They thus interviewed their respondents after testing them to obtain data on the participants' reasoning. In the case of their PCK items ("students and content"), they analyzed these a posteriori think-aloud protocols in terms of whether participants referred explicitly to their knowledge on students' thinking or argued in purely mathematical terms. Interestingly, non-teachers and mathematicians were much more likely to base their response solely on mathematical knowledge. For example, only 1.5% of the mathematicians mentioned students' thinking; in the teacher sample, in contrast, 41% of participants justified their choice by reference to students' thinking.

These findings confirm two issues raised in the context of COACTIV. First, it is difficult to develop items that tap PCK alone, given that CK seems to be one route to PCK. Second, there is another, very teacher-specific route to PCK that is not strictly mathematical. Alonzo (2007) writes: "While some researchers have posited that subject matter knowledge is a pre-requisite for PCK (e.g., van Driel, Verloop, & Vos, 1998), Magnusson, Krajcik, &

Borko (1999) propose multiple pathways to developing PCK: teachers with strong subject matter knowledge *and* those with strong general pedagogical knowledge each build upon their existing knowledge to construct PCK." The latter view is consistent with the findings of Hill et al., (2007) and with the performance of biology/chemistry teachers and NGY teachers on the PCK test in the COACTIV construct validation study.

3.5.2 The MT21 study

The MT21 study investigated three large samples of mathematics teacher candidates in Germany at different stages of their training: at the beginning and the end of the first phase (university) and during the second phase (2-year teaching placement) (Blömeke, Kaiser, & Lehmann, 2008).

How do the MT21 results and the COACTIV results compare? First, a latent correlation of 0.81 was found between PCK and CK in the MT21 sample of 878 teacher candidates, surprisingly close to the latent correlation of 0.79 found in the COACTIV sample (Krauss et al., 2008a). This finding again demonstrates that there seems to be an essential, but not complete, overlap between PCK and CK (which seems to be independent of the details of test conceptualization).

The performances of the three MT21 samples are particularly interesting in the present context. The COACTIV construct validation data suggest a relatively steep increase in teacher knowledge from the beginning to the end of university training, followed by a more modest increase in the second phase, and stagnation after qualification. In the MT21 study, the sample of students at the end of their university training is directly comparable with the corresponding COACTIV sample (sample 5). Relative to COACTIV, however, the three MT21 samples cover a rather shorter period in an aspiring mathematics teacher's career. Whereas the three MT21 samples cover the period from the beginning of university training to teaching practice in schools (second phase of teacher education), the COACTIV analysis extends from grade 13 to practicing teachers. Very interestingly, the MT21 data nevertheless reflect the knowledge growth curve suggested by the COACTIV data; the authors report that the MT21 data also suggest that the great majority of both PCK and CK is acquired at university, with a more modest increase during the second phase of teacher training (for details, see Blömeke, Kaiser, Schwarz et al., 2008, p. 146).

This theoretical and empirical correspondence for German samples in the findings of the COACTIV and MT21 groups (which worked fully independently) not only supports the validity of the underlying constructs of PCK and CK, but also gives reason to hope that key findings might prove to be generally replicable.

4 Summary

In COACTIV, PCK was conceptualized as knowledge of explanations and representations, knowledge of students' thinking, and knowledge of multiple solutions to mathematical tasks. CK was conceptualized as deep background knowledge of school-level mathematics. What can be concluded about the validity of these constructs?

The COACTIV data provided first evidence of the validity of the constructs (Sect. 2.4). For instance, we found differences in CK across school types that were in line with the differences in university training provided for teacher candidates aspiring to teach at the academic track or elsewhere. The lack of positive correlations between teacher knowledge and years of classroom experience may at first seem surprising, but the *deliberate practice* theory of expertise development (Ericsson et al., 1993) provides an explanation for this finding. It must be acknowledged, however, that this finding contradicts other theories that attribute teachers' expertise development to their practical classroom experience (Hashweh, 2005). It is conceivable, however, that routines and automatizations are developed during classroom practice that enable teachers to access and apply their knowledge more rapidly and efficiently (e.g., Hiebert et al., 2002).

External correlations with teachers' subjective beliefs on mathematics and on the learning of mathematics show that knowledgeable teachers reject the views that mathematics is just a toolbox and that mathematics can best be learned by careful listening. As expected, these expert teachers view mathematics rather as a process and believe that it should be learned by means of self-determined active discovery (including reflecting on one's errors, etc.). Moreover, results of structural equation modeling show that PCK, mediated by aspects of the lesson, supports student learning (Baumert et al., 2006, 2008). A solid basis of CK, in turn, appears to facilitate the construction of PCK (Krauss et al., 2008b). These findings are perfectly in line with the theoretical roles usually attributed to CK and PCK.

When contrast populations were administered the COACTIV tests in an extra construct validation study (Sect. 3), all but one of the patterns of results was in accordance with previously formulated hypotheses (the professional knowledge hypothesis and the growing knowledge hypothesis). Grade 13 students enrolled in advanced mathematics courses and biology/chemistry teachers performed poorly on the tests of both CK and PCK; mathematics teacher candidates performed better than the grade 13 students but worse than the COACTIV teachers. Students majoring in mathematics performed expectedly well on the CK items but also surprisingly well on the PCK

items, a finding that may be attributable to several factors (e.g., selectivity of the sample or underrepresentation of geometry items). However, our data did not fully support Shulman's (1987) claim that "pedagogical content knowledge is the category most likely to distinguish the understanding of the content specialist from that of the pedagogue" (p. 8). Although it is important to note in this context that teacher candidates are trained to teach two subjects, these findings may indicate that *very strong* subject matter competence can indeed be one route to pedagogical content knowledge (see also GY teachers). Yet, at the same time, there seems to be another, *teacher-specific* route to PCK: both the NGY teachers and the biology and chemistry teachers in the COACTIV sample attained relatively good PCK scores with poor mathematical CK. Taken together, these findings are in line with Shulman's view of PCK as amalgam of CK and PK. Future research is needed to investigate the specific conditions and processes of their possible mutual compensation.

Remarkably, although there was no collaboration between the COACTIV, MT21, and Michigan groups there is strong theoretical consensus on the key ingredients of pedagogical content knowledge. In particular, either explicitly or implicitly, all groups considered knowledge of explaining the subject and of students' thinking as crucial aspects of PCK. These two ingredients thus seem to be universally accepted as the core of mathematics teachers' pedagogical content knowledge.

Although all three groups could separate PCK and CK empirically, a deep connection between both knowledge categories was found. However, it is conceivable that all three groups' efforts to construct items tapping "pure" PCK have not yet proved successful and that new approaches must be taken to construct PCK items that are not "contaminated" by CK (e.g., by using video clips to increase ecological validity).

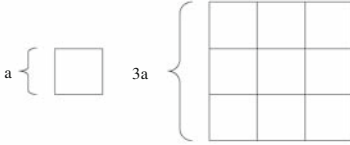
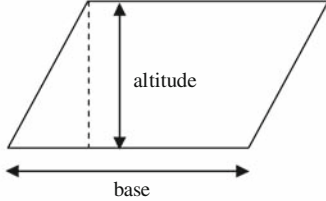
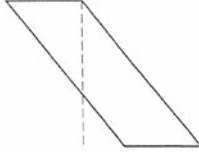
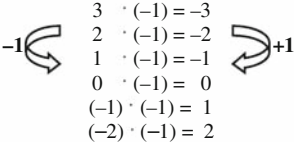
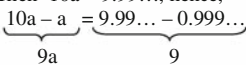
The results for the three samples investigated in the context of the growing knowledge hypothesis are in line with corresponding results from MT21. There seems to be a steep increase in knowledge (both PCK and CK) during university training and then a more gentle increase during the second phase of teacher training (2-year compulsory teaching placement). In sum, conceptual overlap and parallel results indicate the mutual validity of the knowledge constructs developed in the related projects.

However, one varying detail must be mentioned. COACTIV also investigated PCK in terms of knowledge of tasks. Note that this subfacet of PCK is not a hidden facet of CK; in fact, it has the lowest correlation with CK of all three subfacets of PCK (see Krauss et al., 2008b; Krauss et al., 2008a). Rather, this subfacet fits conceptually and psychometrically into COACTIV's PCK approach and can thus be considered theoretical progress. One participant's

exceptional score of 37 on the PCK test (the highest score in our COACTIV sample, which is more than 3 standard deviations above the average) demonstrates the real and substantial scope for improving teacher performance and gives rise to the hope that more teachers can in future be trained to comparable levels, which will substantially benefit their students' learning.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix: Sample items and responses scoring 1 for the COACTIV tests on PCK and CK

Knowledge Category (Subscale)	Sample Item	Sample response (scoring 1)
PCK Task	<p>How does the surface area of a square change when the side length is tripled? Show your reasoning.</p> <p>Please note down as many different ways of solving this problem as possible.</p>	<p><u>Algebraic response</u> 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.</p> <p><u>Geometric response</u> Nine times the area of the original square</p> 
PCK Student	<p>The area of a parallelogram can be calculated by multiplying the length of its base by its altitude.</p>  <p>Please sketch an example of a parallelogram to which students might fail to apply this formula.</p>	 <p>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.</p>
PCK Instruction	<p>A student says: I don't understand why $(-1) \cdot (-1) = 1$</p> <p>Please outline as many different ways as possible of explaining this mathematical fact to your student.</p>	<p>The "permanence principle," although it does not prove the statement, is one way to illustrate the logic behind the multiplication of two negative numbers:</p> 
CK	<p>Is it true that $0.999999\dots = 1$?</p> <p>Please give detailed reasons for your answer.</p>	<p>One possibility: Let $0.999\dots = a$</p> <p>Then $10a = 9.99\dots$, hence, $10a - a = 9.99\dots - 0.999\dots$</p>  <p>Therefore $a = 1$; hence, the statement is true</p>

References

- Ackerman, P. L. (1996). A theory of adult intellectual development: process, personality, interests, and knowledge. *Intelligence*, 22, 227–257.
- Alonzo, A. C. (2007). Challenges of simultaneously defining and measuring knowledge for teaching. *Measurement*, 5(2–3), 131–208.
- Ball, D. L., Hill, H. C., & Bass, H. (2005). Knowing mathematics for teaching. *American Educator*, Fall, 14–46.
- 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 edn., pp. 433–456). New York: Macmillan.
- Barnes, J. (1985). Teaching experience and instruction. In T. Husen & T. N. Postlethwaite (Eds.), *The international encyclopedia of education* (Vol. 9, pp. 5124–5128). Oxford: Pergamon Press.
- Baron-Boldt, J., Schuler, H., & Funke, U. (1988). Prädiktive Validität von Schulabschlußnoten: Eine Meta-Analyse [predictive validity of school leaving grades: a meta-analysis]. *Zeitschrift für pädagogische Psychologie*, 2, 79–90.
- Baumert, J., Blum, W., Brunner, M., Jordan, A., Klusmann, U., Krauss, S., Kunter, M., Neubrand, M., & Tsai, Y.-M. (2006, January). *Teacher knowledge and student progress*. Paper presented at Cubberley Lecture Series, Stanford.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften [Keyword: Teachers' professional competence]. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Dubberke, T., Jordan, A., et al. (2008). *Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress*. (submitted).
- Baumert, J., Kunter, M., Brunner, M., Krauss, S., Blum, W., & Neubrand, M. (2004). Mathematikunterricht aus der Sicht der PISA-Schülerinnen und -Schüler und ihrer Lehrkräfte [Mathematics lessons from the perspective of the PISA students and their teachers]. 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* (S. 314–354). Münster: Waxmann.
- Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, 35(5), 463–482.
- Blömeke, S., Kaiser, G., & Lehmann, R. (Eds.) (2008a). *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare. Erste Ergebnisse zur Wirksamkeit der Lehrerbildung* [Professional competencies of mathematics teacher candidates in Germany. Knowledge, beliefs, and learning opportunities. First findings on the effectiveness of teacher training]. Münster: Waxmann.
- Blömeke, S., Kaiser, G., Schwarz, B., Lehmann, R., Seeber, S., Müller, C., et al. (2008b). Entwicklung des fachbezogenen Wissens in der Lehrerbildung [Development of content knowledge in teacher training]. 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 Lehrerbildung*. Münster: Waxmann.
- Blömeke, S., Lehmann, R., Seeber, S., Schwarz, B., Kaiser, G., Felbrich, A., et al. (2008c). Niveau- und institutsbezogene Modellierungen des fachbezogenen Wissens [Level- and institution-related models of subject-related 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 Lehrerbildung*. Münster: Waxmann.
- Blömeke, S., Seeber, S., Lehmann, R., Kaiser, G., Schwarz, B., Felbrich, A., et al. (2008d). Messung des fachbezogenen Wissens angehender Mathematiklehrkräfte [Measurement of the content knowledge of mathematics teacher candidates]. 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 Lehrerbildung*. Münster: Waxmann.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111, 1061–1071.
- Bromme, R. (1992). *Der Lehrer als Experte: Zur Psychologie des professionellen Wissens [The teacher as an expert: The psychology of professional knowledge]*. Bern: Huber.
- Brunner, M., Kunter, M., Krauss, S., Baumert, J., Blum, W., Neubrand, M., et al. (2006a). Welche Zusammenhänge bestehen zwischen dem fachspezifischen Professionswissen von Mathematiklehrkräften und ihrer Ausbildung sowie beruflichen Fortbildung? [What are the relationships between mathematics teachers' content knowledge and their initial and in-service training?]. *Zeitschrift für Erziehungswissenschaft*, 9(4), 521–544.
- Brunner, M., Kunter, M., Krauss, S., Klusmann, U., Baumert, J., Blum, W., et al. (2006b). Die professionelle Kompetenz von Mathematiklehrkräften: Konzeptualisierung, Erfassung und Bedeutung für den Unterricht; eine Zwischenbilanz des COACTIV-Projekts [The professional competencies of mathematics teachers: Conceptualization, assessment, and significance for instruction; an interim review of the COACTIV project]. In M. Prenzel & L. Allolio-Näcke (Eds.), *Untersuchungen zur Bildungsqualität von Schule: Abschlussbericht des DFG-Schwerpunktprogramms* (pp. 54–82). Münster: Waxmann.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159.
- Cumming, G., & Finch, S. (2005). Inference by eye. Confidence intervals and how to read pictures of data. *American Psychologist*, 60(2), 170–180.
- 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 about teaching and learning: Influences on lessons structure and content and student learning gains]. *Zeitschrift für Pädagogische Psychologie* (in press).
- Educational Testing Service (ETS) (2006). *Proper use of the Praxis series and related assessments*. Retrieved June 19, 2007, from <http://www.ets.org/Media/Tests/PRAXIS/pdf/guidelines.pdf>.
- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100, 363–406.
- Hashweh, M. Z. (2005). Teacher pedagogical constructions: A reconfiguration of pedagogical content knowledge. *Teachers and Teaching: Theory and Practice*, 11(3), 273–292.
- Hiebert, J., Gallimore, R., & Stigler, J. W. (2002). A knowledge base for the teaching profession: What would it look like and how can we get one? *Educational Researcher*, 31(5), 3–15.
- 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, 95–114.

- Hill, H. C., Dean, C., & Goffney, I. M. (2007). Assessing elemental and structural validity: Data from teachers, non-teachers, and mathematicians. *Measurement*, 5(2–3), 81–92.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on students' 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., Krauss, S., Löwen, K., Blum, W., Neubrand, M., Brunner, M., et al. (2008). Aufgaben im COACTIV-Projekt: Zeugnisse des kognitiven Aktivierungspotentials im deutschen Mathematikunterricht [Tasks in the COACTIV project: Testaments to cognitive activation potential in German mathematics instruction]. *Journal für Mathematikdidaktik*, 29(2), 83–107.
- Keeves, J. P. (1992). Longitudinal research methods. In J. P. Keeves (Ed.), *The IEA Technical Handbook* (pp. 15–38). The Hague: IEA.
- Kennedy, M. M., Ball, D. L., & McDiarmid, G. W. (1993). *A study package for examining and tracking changes in teachers' knowledge*. East Lansing, MI: The National Center for Research on Teacher Education/Michigan State University.
- Kirsch, A. (2000). Aspects of simplification in mathematics teaching. In I. Westbury, S. Hopmann & K. Riquarts (Eds.), *Teaching as a reflective practice. The German Didaktik Tradition* (pp. 267–284). Mahwah: Lawrence Erlbaum Associates.
- Klein, F. (1933). *Elementarmathematik vom höheren Standpunkte aus [Elementary mathematics from a higher viewpoint]*. Berlin: Springer.
- Klusmann, U., Kunter, M., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Teachers' occupational well-being and the quality of instruction: The important role of self-regulatory patterns. *Journal of Educational Psychology*, 100(3), 702–715.
- Krauss, S., & Brunner, M. (2008). *Professionelles Reagieren auf Schülerantworten: Ein Reaktionszeittest für Mathematiklehrkräfte* [Professional responses to student answers: A reaction time test for mathematics teachers]. (submitted).
- Krauss, S., Kunter, M., Brunner, M., Baumert, J., Blum, W., Neubrand, M., et al. (2004). COACTIV: Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz [COACTIV: Professional competence of teachers, cognitively activating instruction, and the development of students' mathematical literacy]. In J. Doll & M. Prenzel (Eds.), *Bildungsqualität von Schule: Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung* (pp. 31–53). Münster: Waxmann.
- Krauss, S., Brunner, M., Kunter, M., et al. (2008a). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology*, 100(3), 716–725.
- Krauss, S., Neubrand, M., Blum, W., et al. (im Druck) (2008b). Die Untersuchung des professionellen Wissens deutscher Mathematik-Lehrerinnen und -Lehrer im Rahmen der COACTIV-Studie [Assessment of the professional knowledge of mathematics teachers in the COACTIV study]. *Journal für Mathematikdidaktik (JMD)* 29(3/4) (in press).
- Kunter, M., Brunner, M., Baumert, J., Klusmann, U., Krauss, S., Blum, W., et al. (2005). Der Mathematikunterricht der PISA-Schülerinnen und -Schüler: Schulformunterschiede in der Unterrichtsqualität [The mathematics instruction of the PISA students: Differences in instructional quality across the school types]. *Zeitschrift für Erziehungswissenschaft*, 8(4), 502–520.
- 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 instruction in the PISA classes 2004: Conditions, forms, and processes of teaching learning]. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R., et al. (Eds.), *PISA 2003. Untersuchung zur Kompetenzentwicklung im Verlauf eines Schuljahres* (pp. 161–194). Münster: Waxmann.
- 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: Waxmann.
- Kunter, M., Tsai, Y.-M., Klusmann, U., Brunner, M., Krauss, S., & Baumert, J. (2008). Enjoying teaching. Enthusiasm and instructional behaviors of secondary school mathematics teachers. *Learning and Instruction* (in press).
- Lanahan, L., Scotchmer, M., & McLaughlin M. (2004). *Methodological critique of current NCES survey measures of instructional processes*. Retrieved May 14, 2006, from http://www.air.org/news_events/documents/AERA2004NCESMeasures.pdf.
- Leinhardt, G., & Greeno, J. G. (1986). The cognitive skill of teaching. *Journal of Educational Psychology*, 78(2), 75–95.
- Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler [It depends on the teacher. Empirical evidence for the relations between teacher competence, teacher behavior, and student learning]. In C. Allemann-Ghionda & E. Terhart (Eds.), *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern: Ausbildung und Beruf* (pp. 47–70). Weinheim: Beltz.
- Matz, M. (1982). Towards a process model for high school algebra errors. In D. Sleeman & J. S. Brown (Eds.), *Intelligent tutoring systems* (pp. 25–50). London: Academic Press.
- Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge: The construct and its implications for science education* (pp. 95–132). Dordrecht: Kluwer.
- Mayer, R. E. (2004). Teaching of subject matter. *Annual Review of Psychology*, 55, 715–744.
- Messick, S. (1988). The once and future issues of validity. Assessing the meaning and consequences of measurement. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 33–45). Hillsdale: Lawrence Erlbaum.
- Mieg, H. (2001). *The social psychology of expertise: Case studies in research, professional domains, and expert roles*. Hillsdale: Lawrence Erlbaum.
- Millman, J., Bishop, C. H., & Ebel, R. (1965). An analysis of test-wiseness. *Educational and Psychological Measurement*, 25(3), 707–726.
- Nunally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd edn.). New York: McGraw-Hill.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), 307–322.
- Palmer, D. J., Stough, L. M., Burdinski, T. K., Jr., & Gonzales, M. (2005). Identifying teacher expertise: An examination of researchers' decision making. *Educational Psychologist*, 40(1), 13–25.
- Phelps, G., & Schilling, S. (2004). Developing measures of content knowledge for teaching reading. *Elementary School Journal*, 105, 31–48.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., et al. (2004). *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland: Ergebnisse des zweiten internationalen Vergleichs [PISA 2003. Educational levels in Germany:*

- Findings from the second international comparison*]. Münster: Waxmann.
- 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: Assessment of student progress over the school year]*. Münster: Waxmann.
- Schilling, S. G., & Hill, H. C. (2007). Assessing measures of mathematical knowledge for teaching: A validity argument approach. *Measurement, 5*(2–3), 69–130.
- 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: MSU. Retrieved December 12, 2007, from http://usteds.msu.edu/related_research.asp.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory. A primer*. Newbury Park: Sage.
- Sherin, M. G. (1996). *The nature and dynamics of teacher's content knowledge*. Unpublished doctoral dissertation, University of California, Berkeley.
- 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.
- Silver, E. A., Ghousseini, H., Gosen, D., Charalambous, C., & Font Strawhun, B. T. (2005). Moving from rhetoric to praxis: Issues faced by teachers in having students consider multiple solutions for problems in the mathematics classroom. *Journal of Mathematical Behavior, 24*, 287–301.
- Sternberg, R. J., & Horvath, J. A. (1995). A prototype view of expert teaching. *Educational Researcher, 24*(6), 9–17.
- Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Peck, R., & Rowley, G. (2008). *Teacher Education and Development Study in Mathematics (TEDS-M)*. East Lansing, MI: Teacher Education and Development International Study Center, College of Education, Michigan State University.
- van Driel, J. H., Verloop, N., & de Vos, W. (1998). Developing science teachers' pedagogical content knowledge. *Journal of Research in Science Teaching, 35*, 673–695.