Journal of Educational Psychology
2008, Vol. 100, No. 3, 716–725

0022-0663/08/$12.00 DOI: 10.1037/0022-0663.100.3.716

Pedagogical Content Knowledge and Content Knowledge of Secondary Mathematics Teachers

Stefan Krauss, Martin Brunner, Mareike Kunter, and Jürgen Baumert
Max Planck Institute for Human Development

Werner Blum
University of Kassel

Michael Neubrand
University of Oldenburg

Alexander Jordan
University of Bielefeld

Drawing on the work of L. S. Shulman (1986), the authors present a conceptualization of the pedagogical content knowledge and content knowledge of secondary-level mathematics teachers. They describe the theory-based construction of tests to assess these knowledge categories and the implementation of these tests in a sample of German mathematics teachers ($N = 198$). Analyses investigate whether pedagogical content knowledge and content knowledge can be distinguished empirically, and whether the mean level of knowledge and the degree of connectedness between the two knowledge categories depends on mathematical expertise. Findings show that mathematics teachers with an in-depth mathematical training (i.e., teachers qualified to teach at the academic-track Gymnasium) outscore teachers from other school types on both knowledge categories and exhibit a higher degree of cognitive connectedness between the two knowledge categories.

Keywords: professional knowledge of mathematics teachers, pedagogical content knowledge, content knowledge, expert teachers

Imagine you are a mathematics teacher. A student puts his hand up and says: "I don't understand why −1 times −1 equals +1. I know it's the correct result, but it makes no sense to me. Why does multiplying two negative numbers give a positive number?" How would you explain this result to your student? Scenarios like these are typical for the task of teaching. In order to respond appropriately, teachers not only need to understand the mathematical concepts underlying the question, they also need to know how these concepts can best be explained to students.

The relevance of teachers' domain-specific knowledge to high-quality instruction has been discussed, particularly in the context of mathematics teaching (Ball, Lubienski, & Mewborn, 2001; Fennema & Franke, 1992). Drawing mostly on qualitative data, it has been shown that a deep understanding of mathematical concepts may enable teachers to access a broad repertoire of strategies for explaining and representing mathematical content to their students (Ball, Hill, & Bass, 2005; Ma, 1999). First quantitative evidence shows that students' learning gains in mathematics may be predicted by their teachers' mathematics-related knowledge (Hill, Rowan, & Ball, 2005). The precise nature of teachers' knowledge (i.e., the content and structure of knowledge relating to specific school subjects) remains empirically uncertain, however. Following Shulman (1986, 1987), a theoretical distinction is often drawn between domain-specific subject-matter knowledge, content knowledge (CK), and the knowledge needed for teaching a specific subject, pedagogical content knowledge (PCK). These two knowledge categories may be hypothesized to represent conceptually distinct forms of knowledge, with the former perhaps being the prerequisite for the development of the latter. Alternatively, it is conceivable that the two are merged to form a single body of domain-specific knowledge for teaching (for first results on the knowledge of elementary teachers, see Hill, Schilling, & Ball, 2004, or Phelps & Schilling, 2004). Despite its great relevance to the development of teachers' knowledge and possible implications for teacher training curricula, this issue remains empirically unresolved, primarily because very few instruments are yet available to tap teachers' knowledge directly.

In this article, we present an empirical approach to assessing the knowledge of secondary-level mathematics teachers and investigate whether the theoretical distinction between pedagogical content knowledge and content knowledge can be verified empirically. We report on the theory-driven construction of a test to assess the

Stefan Krauss, Martin Brunner, Mareike Kunter, Jürgen Baumert, Center for Educational Research, Max Planck Institute for Human Development, Berlin, Germany; Werner Blum, Department of Mathematics, University of Kassel, Germany; Michael Neubrand, Department of Mathematics, University of Oldenburg, Germany; Alexander Jordan, Department of Mathematics, University of Bielefeld, Germany.

Martin Brunner is now at the Educational Measurement and Applied Cognitive Science (EMACS) Research Unit, University of Luxembourg.

This research, which was part of the larger COACTIV study (see http://www.mpibberlin.mpg.de/coactiv/index.htm), was funded by the German Research Foundation (DFG). We thank our colleagues at the Center for Educational Research for their invaluable support with this research project and their comments on previous versions of this manuscript, and Susannah Goss for English editing. We also thank the participating teachers and raters for their support.

Correspondence concerning this article should be addressed to Stefan Krauss, who is now at FB 17, Department of Mathematics, University of Kassel, 34109, Kassel, Germany. E-mail: skrauss@mathematik.uni-kassel.de

knowledge categories of PCK and CK directly, and on its application in a German teacher sample. Our investigation of commonalities and differences in CK and PCK was informed by the expertise research literature, which has repeatedly found that the knowledge base of experts is not only more extensive than that of novices, but also more connected and integrated (Chi, Feltovich, & Glaser, 1981; Schmidt & Boshuizen, 1992; Simon & Chase, 1973). We therefore investigated whether the association between PCK and CK varies according to the level of mathematical expertise, hypothesizing that the association between the two categories of knowledge will be substantially closer in teachers high in mathematical expertise, possibly even constituting one inseparable body of knowledge. We addressed this question by capitalizing on a quasi-experimental situation that is specific to teacher education in Germany. The amount (and depth) of content-specific instruction provided to teacher candidates varies substantially depending on whether or not they aspire to teach in the academic-track Gymnasium (GY). One group of teachers can therefore be identified as "mathematical content experts."

The Concepts of Pedagogical Content Knowledge and Content Knowledge

The distinction between teachers' knowledge about teaching in a specific domain (PCK) and their domain-specific subject-matter knowledge (CK) has been widely embraced by the research community. The core meaning of pedagogical content knowledge is best represented by Shulman's (1986, pp. 9–10) original definition, which states that pedagogical content knowledge includes knowledge on how best to represent and formulate the subject to make it comprehensible to others, as well as knowledge on students' subject-specific conceptions and misconceptions (see also Grossman, 1990). Content knowledge, on the other hand, describes a teacher's understanding of the structures of his or her domain. According to Shulman, "The teacher need not only understand *that* something is so, the teacher must further understand *why* it is so" (Shulman, 1986, p. 9), which implies that teachers' content knowledge should represent a deep understanding of the material to be mastered by the students. Shulman's definitions describe teachers' pedagogical content knowledge and content knowledge in generic terms. However, domain-specific approaches have been found to provide more valuable, in-depth insights into instructional processes and their prerequisites (Mayer, 2004b).

Mathematics Teachers' Pedagogical Content Knowledge and Content Knowledge

Research has identified several aspects that are specifically important to successful mathematics instruction, and that might therefore be used to conceptualize pedagogical content knowledge in a mathematics-specific approach. Most importantly, tasks play a central role in mathematics instruction (Christiansen & Walther, 1986), accounting for much of the time allocated to mathematics lessons. Appropriately selected and implemented mathematical tasks lay the foundations for students' construction of knowledge and represent powerful learning opportunities (de Corte, Greer, & Verschaffel, 1996; Williams, 2002). Knowledge about the potential of mathematical tasks for learning is thus a first important aspect of mathematical pedagogical content knowledge. Second, teachers need to work with students' existing conceptions and prior knowledge. Because errors and 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 comprehension difficulties. Students' construction of knowledge often only succeeds with instructional support and guidance (Mayer, 2004a), which may entail various forms of explanations or the explicit use of representations. The knowledge of appropriate mathematics-specific instructional methods is thus a third important component of mathematical pedagogical content knowledge. Whereas the latter two components are based directly on Shulman's (1986) generic conceptualization, we added the component of knowledge on tasks as a third, mathematics-specific component of mathematical pedagogical content knowledge.

The domain-specific conceptualization of teachers' content knowledge seems straightforward. Clearly, teacher knowledge should go beyond an awareness of the material to be mastered by students; rather, teachers should possess mathematical background knowledge of the content covered in the school curriculum at a much deeper level of understanding than their students. This background knowledge of school curriculum content forms a knowledge base that is specific to teachers, in that it overlaps only partially with the mathematics typically taught at university.

The Interplay Between Pedagogical Content Knowledge and Content Knowledge

Although the distinction between pedagogical content knowledge and content knowledge appears highly plausible at a theoretical level, its empirical basis is far from certain. Despite repeated calls for reliable and valid measurement assessments of teachers' knowledge (e.g., Lanahan, Scotchmer, & McLaughlin, 2004), instruments suitable for assessing the categories of teachers' knowledge remain scarce. To date, only a few studies have investigated PCK and CK empirically. In some of these studies, mathematics teachers' content knowledge was not tested explicitly, because researchers did not want to give the impression of "testing" teachers (e.g., Kennedy, Ball, & McDiarmid, 1993). Other studies aiming to tap both knowledge categories found that, although it is possible to construct separate tests in principle, the overlap between the two categories was in fact so high that a global factor of knowledge relating to mathematics instruction seemed just as likely (Hill et al., 2004; Kahan, Cooper, & Betha, 2003). At the same time, a study with elementary teachers in the domain of reading found separate categories only, and no common factor (Phelps & Schilling, 2004). One explanation for these inconclusive results may be that the structure of knowledge differs across teacher populations. Studies comparing the knowledge base of experienced and novice teachers (for an overview see Berliner, 2001) suggest that expert teachers not only know more than novice teachers, but that their knowledge is differently structured and may be more highly integrated. This conclusion is in line with findings from expertise research in other domains, which show that experts' knowledge bases are usually not only more extensive than those of novices, but also more connected and integrated (Chi et al., 1981; Schmidt & Boshuizen, 1992; Simon & Chase, 1973). Whether or not teachers' pedagogical content knowledge and content knowledge are separable categories of knowledge may therefore be a function of different levels of expertise. All previous empirical attempts to test models of teachers' knowledge have looked at elementary teachers, who can be assumed to have lower levels of subject-specific expertise. It thus seems worth examining the distinction between CK and PCK in a sample of secondary-level teachers.

The Present Investigation

The goal of the present article was to construct and to establish a test of secondary mathematics teachers' PCK and CK and to use this test to examine the level and the connectedness of the two knowledge categories in two groups of teachers with different mathematical expertise. To specify our research hypotheses and make them empirically testable, we first clarified our approach to the concepts of *connectedness* and *mathematical expertise*.

Our methodological approach is based on a structural equation framework, in which PCK and CK are conceptualized as latent variables. The latent correlation between PCK and CK is particularly relevant to the issue of distinguishability (see Figure 1a). A higher latent correlation between PCK and CK indicates higher cognitive connectedness between the two knowledge categories. Moreover, a correlation close to one may indicate that the two knowledge constructs indeed form a single, indistinguishable body of knowledge on the cognitive level. In the PCK–CK model (Figure 1a), the means of the two knowledge constructs are depicted by paths from the triangle representing the mean structure to the two latent variables.

We used a quasi-experimental approach to examine whether higher connectedness is a function of higher expertise, attributing expert status on the basis of the teachers' university training (for an overview of alternative ways to identify expert teachers, see Palmer, Stough, Burdenski, & Gonzales, 2005). In Germany, all candidates entering a teacher training program must have graduated from the highest track in the school system, the GY, and received the Abitur qualification. At university, those aspiring to teach at the secondary level must choose between separate degree programs qualifying them to teach either at GY or in the other secondary tracks (e.g., Realschule or Sekundarschule). GY and non-Gymnasium (NGY) teacher education students are usually strictly separated during their university training. One of the main differences in their degree programs is the subject matter covered: Students training to teach at GY cover an in-depth curriculum comparable to that of a master's degree. Relative to their colleagues, who receive less subject-matter training (and usually spend less time at university), GY teachers may therefore be considered mathematical experts, and the two groups of teachers may be contrasted in a quasi-experimental approach. Given that previous studies (e.g., Hill et al., 2004) have presented evidence for a close connection between the two subject-specific knowledge categories, and because CK is often discussed as a prerequisite for PCK, we expected GY teachers would score higher in the PCK test as well, although the difference would probably not be as pronounced as for the CK test.

On the basis of this conceptualization of connectedness and mathematical expertise, we formulated our research hypotheses as follows:

> *Hypothesis 1:* GY teachers significantly outscore their colleagues from NGY school types on CK and PCK.

> *Hypothesis 2:* The latent correlation between PCK and CK is substantially higher for GY teachers than for NGY teachers (perhaps even approaching $r = 1$).

Method

Participants

The present analyses are based on data obtained from 198 secondary mathematics teachers in Germany. Participants taught mathematics in 10th-grade classes sampled within the framework of a nationally representative student achievement study (cf. Kunter et al., 2007). Thus, our teacher sample can be considered fairly representative of 10th-grade mathematics teachers in Germany. Of the 198 teachers, 85 (55% male) taught at the academic-track GY, and 113 (43% male) at other secondary school types (NGY; e.g., Realschule, Sekundarschule). The average age of participating teachers was 47.2 years ($SD = 8.4$); 46.4 years ($SD = 9.1$; range 28–65) in the GY group and 47.8 years in the NGY group ($SD = 7.7$; range 28–62). Teachers were paid 60 Euro (approximately US $60 at the time of the survey) for their participation.

Assessment of PCK and CK

PCK test. In line with our theoretical framework, the PCK test contained three subscales: knowledge of mathematical tasks (Task), knowledge of student misconceptions and difficulties (Student), and knowledge of mathematics-specific instructional strategies (Instruction). Because the potential of tasks for students' learning can be exploited by considering various solution paths (e.g., Silver, Ghousseini, Gosen, Charalambous, & Strawhun, 2005), we assessed knowledge of tasks by testing teachers' awareness of multiple solution paths: four items required teachers to list as many different ways of solving the task as possible. Knowledge of student misconceptions and difficulties was assessed by presenting teachers with seven scenarios and asking them to detect, analyze (e.g., give cognitive reasons for a comprehension problem), or predict a typical student error or a particular comprehension difficulty. Knowledge of subject-specific instructional strategies was assessed by 10 items requiring teachers to explain mathematical situations (e.g., to provide useful representations, analogies, illustrations, or examples to make mathematical content accessible to students). Sample items for the three PCK subscales (Task, Student, Instruction) are displayed in Figure 2, along with sample responses scoring 1.

CK test. We conceptualized content knowledge as in-depth background knowledge on the contents of the secondary-level mathematics curriculum. Thirteen items were constructed to cover relevant content areas (e.g., arithmetic, algebra, and geometry) and to tap conceptual or procedural skills (Figure 2 presents a sample item).

Scoring scheme. All 34 items assessing PCK and CK were open ended. Items with no response or an incorrect response were scored 0; each correct answer was scored 1 (for items requiring several answers, e.g., the multiple solution tasks, the sum of the correct answers was calculated). Each test item was coded by two trained raters independently; in the event of rater disagreement, consensus was reached through discussion. The interrater reliabil-

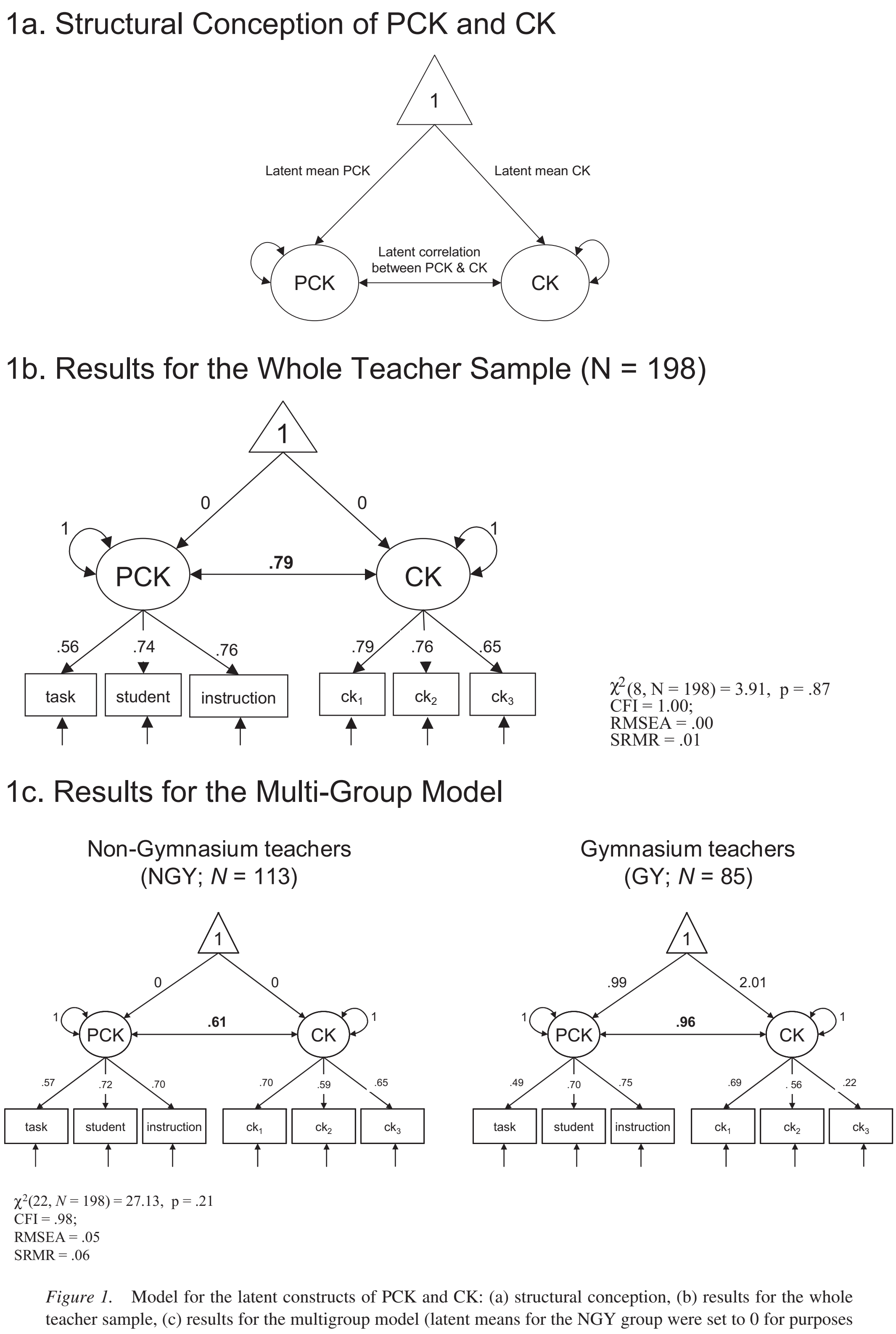


Figure 1. Model for the latent constructs of PCK and CK: (a) structural conception, (b) results for the whole teacher sample, (c) results for the multigroup model (latent means for the NGY group were set to 0 for purposes of model identification. Model fit indices and standardized model parameters are shown for (b) and (c). SRMR values below .08, RMSEA values below .05, and CFI values above .95 can be considered indicative of a good model fit. PCK = pedagogical content knowledge; CK = content knowledge; GY = Gymnasium; NGY = non-Gymnasium; SRMR = standardized root mean residual; RMSEA = root-mean-square error of approximation; CFI = comparative fit index.

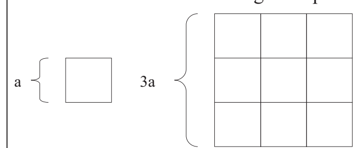
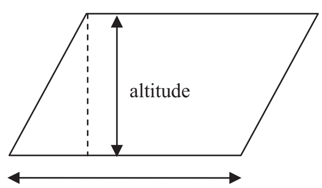
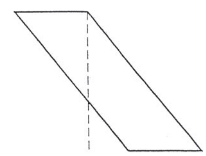
Knowledge Category (Subscale)	Sample Item	Sample response (scoring 1)
PCK Task	How does the surface area of a square change when the side length is tripled? Show your reasoning. Please note down as many different ways of solving this problem (and different reasonings) as possible.	Algebraic response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

Figure 2. Sample items and corresponding sample responses (scoring 1) from the three subscales of the pedagogical content knowledge (PCK) test (Task, Student, Instruction) and from the content knowledge (CK) test.

ity ρ (Shavelson & Webb, 1991) was satisfactory (on average, ρ was .81 with *SD* = .17).

Procedure. The assessment of PCK and CK was conducted individually in a separate room at the teacher's school on a workday afternoon. It was administered by a trained test administrator, as a power test with no time constraints. The average time required to complete the 34 items was about 2 hr (approx. 65 min for the PCK instrument and 55 min for the CK instrument). The teachers were not allowed to use calculators.

Statistical Analysis of the PCK–CK Measurement Model

To investigate the structure of knowledge, we employed a confirmatory factor analysis in which the two knowledge categories were conceptualized as latent constructs based on manifest indicators. The sample size required for structural equation modeling in general and for confirmatory factor analysis in particular has been a matter of some debate. Some scholars recommend a sample size of at least $N = 200$; others argue that no general recommendations can be made because sample size depends strongly on the properties of the model investigated and the data to be analyzed (cf. Jackson, 2003; Marsh, Hau, Balla, & Grayson, 1998). Another way of approaching the issue is to ensure a certain ratio of estimated parameters to participants. Some authors suggest a ratio of at least 1:5 (cf. Marsh et al., 1998). Given the 34 items in our tests (21 measuring PCK and 13 measuring CK), a PCK–CK (item factor) model in which each of the constructs is measured by the respective items would give a ratio of about 1:3, which is unacceptable.

To guarantee that the conclusions derived from our models are valid given our relatively small sample size of $N = 198$ teachers, we therefore used subscale scores and parcel scores rather than items as manifest indicators (Little, Cunningham, Shahar, & Widaman, 2002), thus reducing the number of parameters to be estimated (e.g., instead of 34, there are just 6 factor loadings; see Figure 1b). This reduction in parameter numbers is particularly relevant for the second part of the analysis, in which the sample is divided into two groups. The latent knowledge construct PCK was measured by the three subscale scores, which were calculated by summing the corresponding item scores. The latent knowledge construct CK was measured by three parcel scores (Little et al., 2002). A preliminary exploratory factor analysis did not identify specific and interpretable subdimensions of CK, but showed that the CK items form one single factor. Therefore, four items with the same stem were assigned to one parcel (ck_3) to account for potential task-specific variance that is represented in the corresponding residual term. The remaining nine (unidimensional) items were randomly divided into two further parcels (ck_1 and ck_2) following Little et al. (2002).

To tackle the question of empirical distinguishability, we first analyzed the PCK–CK model across all teachers (see Figure 1b), investigating the latent correlation between the two knowledge categories. We then specified a multigroup model (see Figure 1c) to address differences between the two groups of teachers. Although our sample size of $N = 198$ teachers was sufficient to analyze the PCK–CK model across all teachers (yielding a ratio of about 1:10; see Figure 1b), the multigroup model (Figure 1c) requires closer inspection (here the ratio is about 1:6). Although a ratio of 1:6 seems reasonable, the sample size for the multigroup model is at the lower limit. We therefore investigated the power of the statistical analyses conducted. Table 1 lists the four null hypotheses, H_{01}–H_{04}, that were central to our statistical analyses. H_{01} is the null hypothesis stating that PCK and CK do not constitute separable knowledge categories and are not distinguishable in the whole teacher sample. H_{02} and H_{03} are the null hypotheses corresponding to our research Hypothesis 1, and H_{04} is the null hypothesis corresponding to our research Hypothesis 2. Whereas H_{01} can be seen as a restriction on the model across all teachers (see Figure 1b), H_{02}–H_{04} can be seen as particular restrictions on the multigroup model (see Figure 1c). Power can be defined as the probability to reject the null hypothesis when it is wrong. Taking this perspective, we followed the methodological recommendations of Satorra and Saris (1985) and calculated within the latent variable framework the power to reject the null hypotheses H_{01} – H_{04} given our sample size of $N = 198$. To this end, we specified "null models" for H_{01}–H_{04} (e.g., for H_{01}, the latent correlation in Figure 1b, was set to 1). The chi-square difference test statistic that results from subtracting the chi-square test statistic of the original model (e.g., the latent correlation in Figure 1b is .79) from the chi-square test statistic of the corresponding "null model" represents the noncentrality parameter that is necessary to calculate statistical power from a noncentral chi-square distribution according to Satorra and Saris (1985). As shown in Table 1, the statistical power to test our hypotheses was acceptable with a sample size of $N = 198$.

Table 1
Power Analysis ($N = 198$; $\alpha = .05$)

Null hypotheses to reject	$\Delta\chi^2$	Δdf	Power
H_{01} latent correlation $r_{PCK,\ CK}$ is 1 in the total teacher sample	20.57	1	0.99
H_{02} latent mean of PCK is identical in both groups of teachers	37.80	1	0.99
H_{03} latent mean of CK is identical in both groups of teachers	106.28	1	1.00
H_{04} latent correlation $r_{PCK,\ CK}$ is identical in both groups of teachers	6.91	1	0.75

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

Before comparing the latent means and correlations of two groups, we had to investigate whether the constructs measured have the same meaning in both groups. In the structural equation literature, it is standard procedure to run a series of invariance tests when doing multigroup comparisons to establish measurement equivalence by showing that (a) a model with factor loadings and intercepts of the manifest variables constrained to be equal across groups fits reasonably well (scalar invariance) and (b) the fit of the scalar invariant model is not much worse than that of a configural invariant or a metric invariant model that imposes fewer equality constraints across groups (Little, 1997; Vandenberg & Lance, 2000). In a configural invariant model, the factorial structure is assumed to be the same across groups; however, factor loadings and intercepts may differ. In the metric invariant model, the factorial structure and factor loadings are assumed to be identical across groups; however, the intercepts may vary. When local misfit is identified in one of these models, the corresponding equality constraint can be relaxed, and a partial invariant model may allow the same conclusions to be drawn as the full invariant model (Byrne, Shavelson, & Muthén, 1989). Our results (see Table 2) show that the fit of the metric invariant model was considerably worse than that of the configural invariant model. However, freeing the factor loading of ck_3 in group GY led to a substantial improvement in model fit (partial metric invariance). The fit of the partial scalar invariant model (without equality constraint on the factor loading of ck_3) was worse than that of the partial metric invariant model. The cause of the misfit was the invariance constraint on the intercept of ck_3. When this intercept was freely estimated in the GY group, the fit of the partial invariant model

Table 2
Series of Models Investigating Measurement Equivalence for the Multigroup Model

Model	χ^2	*df*	*p*	CFI	RMSEA	SRMR	$\Delta\chi^2$	Δdf
M1. Configural invariance	15.79	16	.47	1.00	.00	.04		
M2. Metric invariance	36.12	20	.01	0.93	.09	.10		
Difference between M1 and M2							24.14***	4
M3. Partial metric invariance[a]	19.75	19	.41	1.00	.02	.05		
Difference between M2 and M3							57.99***	1
M4. Partial scalar invariance[a]	41.70	23	.01	0.92	.09	.10		
Difference between M4 and M3							25.56***	4
M5. Partial scalar invariance[b]	27.13	22	.21	0.98	.05	.06		
Difference between M5 and M4							33.82***	1

Note. Values for $\Delta\chi^2$ were calculated according to the formulas provided by B. O. Muthén (1998-2004, p. 22), which generate corrected $\Delta\chi^2$ statistics when the maximum likelihood estimator MLM is used. CFI = comparative fit index; RMSEA = root-mean-square error of approximation; SRMR = standardized root mean residual.
[a] Factor loading of ck_3 in group Gymnasium freely estimated. [b] Factor loading and intercept of ck_3 in group Gymnasium freely estimated.
*** $p < .001$.

improved and became acceptable (see Table 2). According to Byrne et al. (1989), the partial scalar invariant model, which is presented in Figure 1c, guarantees measurement equivalence; the model can therefore be used to investigate our primary research questions. Furthermore, additional power analyses showed that our sample size yielded a statistical power of at least .99 to distinguish between the models reported in Table 2.

Analyses were run with the Mplus 4.2 software package (L. K. Muthén & Muthén, 1998–2006). Because prior analyses indicated that the distributions of five of the six sum scores (all but the subtest Instruction) deviated from normal distributions (skewness ranged from −.65 to .82; kurtosis ranged from −1.09 to .07), parameters were computed using the robust maximum likelihood estimator MLM, which is recommended for nonnormally distributed data (L. K. Muthén & Muthén, 1998–2006).

Results

Before we address our core hypotheses on the level and connectedness of PCK and CK in the two groups of teachers, we report the psychometric properties of our test scales and examine the general distinguishability of the two knowledge categories in the whole sample of teachers.

Psychometric Properties of PCK and CK and Descriptive Information

To give a first impression of measurement quality, we report the psychometric properties of the overall PCK (21 items) and CK (13 items) scales, which were analyzed by means of parameters derived from classical test theory. Both overall scale scores showed satisfactory reliability, with a Cronbach's alpha of $\alpha = .77$ for the PCK scale and $\alpha = .83$ for the CK scale. The items forming each scale discriminated adequately as evident from their (part-whole corrected) item-total correlations (PCK: $M = .33$, range: .17 to .45; CK: $M = .48$; range: .30 to .66). The means and standard deviations for the subscale scores of PCK and the parcel scores of CK are given in Tables 3 and 4, along with the respective intercorrelations. Table 3 shows these descriptives for the whole teacher sample; Table 4, for the GY and NGY groups separately. In sum, these analyses indicate that the newly developed tests succeeded in reliably assessing the knowledge categories PCK and CK in secondary mathematics teachers.

Distinguishability of PCK and CK

Figure 1b presents the results (parameter estimations and fit indices) for the PCK–CK model across all teachers. The fit of the model, as well as the valence and size of the factor loadings, indicates that the structure of teachers' professional knowledge is accurately captured by two latent constructs representing PCK and CK. The latent correlation between PCK and CK was rather high, at .79, indicating that teachers with higher CK also tend to have higher PCK. To test statistically whether CK and PCK are distinguishable constructs, we estimated a model in which the correlation between CK and PCK was constrained to 1. This led to a statistically significant decline in model fit, as shown by the chi-square difference test, $\Delta\chi^2(1, N = 198) = 20.57$, $p < .001$. Hence, from the perspective of inferential statistics, it can be concluded that, despite their rather high correlational interdependence, CK and PCK represent different constructs, at least when all teachers are considered together.

Table 3
Measures of Teachers' Professional Knowledge (Subscales and Parcels): Correlations and Descriptive Statistics for the Total Teacher Sample

Correlations	Task	Student	Instruction	ck_1	ck_2	ck_3
PCK						
Task (4 items)						
Student (7 items)	.43					
Instruction (10 items)	.43	.56				
CK						
ck_1 (4 items)	.33	.48	.45			
ck_2 (5 items)	.30	.44	.50	.60		
ck_3 (4 items)	.29	.39	.40	.52	.48	
M	6.96	4.93	7.87	1.89	1.47	2.58
SD	2.16	1.98	3.34	1.32	1.27	1.50

Note. Task, Student, and Instruction are the subscales of pedagogical content knowledge (PCK); ck_1, ck_2, and ck_3 are the parcels of content knowledge (CK).

Table 4
Measures of Teachers' Professional Knowledge (Subscales and Parcels): Correlations and Descriptive Statistics by Teacher Group

Correlations	Task	Student	Instruction	ck_1	ck_2	ck_3
PCK						
Task		.28	.39	.34	.30	.11
Student	.48		.53	.44	.26	.11
Instruction	.40	.47		.46	.55	.24
CK						
ck_1	.22	.23	.23		.43	.04
ck_2	.18	.36	.30	.46		.21
ck_3	.30	.29	.33	.44	.38	
NGY						
M	6.60	4.27	6.96	1.18	0.96	1.88
SD	2.02	1.85	3.16	0.97	1.02	1.50
GY						
M	7.45	5.81	9.08	2.84	2.16	3.51
SD	2.27	1.80	3.19	1.11	1.23	0.88
Cohen's *d* (mean GY vs. NGY)	0.40	0.85	0.67	1.60	1.08	1.28

Note. The correlations for Gymnasium (GY) and non-Gymnasium (NGY) teachers are above and below the diagonal, respectively; Task, Student, and Instruction are the subscales of pedagogical content knowledge (PCK); ck_1, ck_2, and ck_3 are the parcels of content knowledge (CK); positive *d* values indicate that GY teachers outperformed NGY teachers. According to Cohen (1992), $d = .3$ represents a small effect, $d = .5$ represents a medium effect, and $d = .8$ represents a large effect. All mean differences are significant at $p < .01$.

Level of Knowledge and Cognitive Connectedness in the Two Groups of Teachers

As expected, GY teachers and NGY teachers differed substantially in their mean levels of knowledge (Table 4 gives the means, standard deviations, and intercorrelations of the manifest measures for each teacher group separately). As indicated by Cohen's *d*, the group differences were particularly pronounced in the CK parcels, substantiating our assumption that GY teachers show much deeper mathematical understanding. This finding was further corroborated by the results of the multigroup model, in which PCK and CK were estimated as latent constructs in both groups simultaneously, thus eliminating measurement error (Figure 1c). The results show that GY teachers clearly have a more extensive knowledge base than NGY teachers in both knowledge categories. The latent means of the GY group (PCK: .99, CK: 2.01) were substantially higher than those of the NGY group (fixed to 0 for both constructs; cf. Little, 1997). The corresponding effect sizes were $d = 1.00$ for latent PCK and $d = 2.15$ for latent CK. These differences were statistically significant, as confirmed by two separate chi-square difference tests between the model depicted in Figure 1c and models in which the latent means of CK or PCK, respectively, were constrained to be the same (i.e., 0) across groups (CK: $\Delta\chi^2[1, N = 198] = 106.28$, $p < .001$; PCK: $\Delta\chi^2[1, N = 198] = 37.80$, $p < .001$), thus corroborating research Hypothesis 1.

Most important, the multigroup model showed that the two teacher groups differed not only in their knowledge level, but also in the structure of their knowledge base. As Figure 1c shows, the findings also supported research Hypothesis 2, that cognitive connectedness is dependent on the level of mathematical expertise: the latent correlation between CK and PCK was .96 in the GY group and .61 in the NGY group. The statistical significance of the difference between these correlations was again confirmed in a chi-square difference test, $\Delta\chi^2(1, N = 198) = 6.91$, $p < .05$, thus showing a substantially higher degree of cognitive connectedness between the two knowledge categories for teachers in the GY group.

Given the extremely high correlation between PCK and CK in the group of GY teachers, the question arises of whether Shulman's (1986, 1987) two subject-specific knowledge categories are in fact empirically distinguishable in this group of highly knowledgeable teachers. To address this question, we compared the multigroup model (Figure 1c) with a model in which the correlation between the two constructs was fixed to 1 for the GY group. The chi-square difference test between the two nested models was not significant, $\Delta\chi^2(1, N = 198) = 0.14$, $p = .72$, indicating that the two knowledge categories form one body of connected knowledge and that PCK and CK are almost indistinguishable in the group of GY teachers. By contrast, when the correlation between the two constructs was fixed to 1 for the NGY group, the chi-square difference was significant, $\Delta\chi^2(1, N = 198) = 20.38$, $p < .001$, indicating that the two knowledge categories are separate in the group of NGY teachers.

The results thus demonstrate that the two groups of teachers, whose university education differed substantially, differ in both the level and structure of their knowledge. However, we cannot yet rule out the possibility that this finding is simply a manifestation of differences that existed between the groups prior to their teacher training. Although both teacher training tracks have the same formal requirements (Abitur qualification, no entrance exams), it is reasonable to assume that higher achieving students tend to opt to teach at Gymnasium. We investigated this possibility by looking at the teachers' own Abitur grade (which corresponds to the U.S. grade-point average), which was assessed in a biographical questionnaire (the Abitur grade was *z*-standardized for our analyses, with higher values indicating a better grade). The GY teachers indeed had substantially higher Abitur grades than the NGY teachers (NGY: $M = -.22$, $SD = 1.03$; GY: $M = .26$, $SD = .92$; $d = .49$, $p < .001$). To test whether this difference might explain the differences found between the two teacher groups to some degree (Figure 1c), we reran the analyses, with mean differences in the Abitur grade between the two groups of teachers being partialed out of the manifest indicators (subscale scores and parcel scores). The latent means and correlations obtained in the multigroup model were thus adjusted for the mean differences in the Abitur grade. The findings confirmed the pattern of results from our previous analyses: the mean differences found between the two teacher groups were of similar magnitude (standardized latent means of the GY teachers (PCK: .88, $d = .88$; CK: 1.79, $d = 1.95$), as were the latent correlations between PCK and CK (GY: $r = .94$, NGY: $r = .60$). Although the Abitur grade is just one indicator of prior differences, these analyses suggest that, even given the same levels of prior knowledge, it is only in GY teachers that pedagogical content knowledge and content knowledge fuse to form one inseparable category of knowledge.

Discussion

In previous studies, most conclusions about the nature of teachers' knowledge have been drawn using indicators that are rather distal to the concept, such as university grades, number of subject matter courses taken at university (cf. Hill et al., 2005), or questionnaire data on beliefs or subjective theories (cf. Pajares, 1992). Consequently, numerous calls have been made in the literature for more valid and reliable assessments of teacher knowledge (e.g., Lanahan et al., 2004). In the present study, we constructed and implemented tests to assess the pedagogical content knowledge and content knowledge of secondary mathematics teachers directly. We took a subject-specific approach, thus responding to repeated calls in the literature for general educational psychological theories to be specified for specific school subjects (e.g., Mayer, 2004b); here, for mathematics. Both knowledge categories were measured reliably, the fit for a corresponding structural model was satisfactory, and the mean differences between teachers with different educational backgrounds provided evidence for the empirical validity of the tests.

Our findings provide further evidence for the applicability of Shulman's (1986, 1987) taxonomy of teacher knowledge in empirical settings. More specifically, being informed by literature on expertise research, they offer a possible interpretation for previous inconclusive findings on the distinguishability of CK and PCK (Hill et al., 2004; Kahan et al., 2003; Phelps & Schilling, 2004). Our findings show that the degree of cognitive connectedness between PCK and CK in secondary mathematics teachers is a function of the degree of mathematical expertise. Capitalizing on the quasi-experimental situation of teacher training in Germany, which allowed us to identify teachers with different levels of mathematical expertise, we found that it was not possible to distinguish the two knowledge categories empirically in the high-expertise group of GY teachers, but that this distinction was clearly visible in the group of NGY teachers. Our findings are thus consistent with findings from other domains of expertise research showing that higher expertise often involves stronger integration of different knowledge categories, or "encapsulated knowledge" (Schmidt & Boshuizen, 1992). Consequently, subject-specific knowledge seems to form a common body of expertise in GY teachers, with high levels of CK and PCK alike. Given that GY teacher candidates receive additional in-depth mathematics training at university, but no additional training in teaching mathematics (relative to NGY teacher candidates), their substantially higher PCK scores are remarkable, and may—although very tentatively—be interpreted as a first indication that CK supports the development of PCK.

Practically, our results have at least two implications. First, our instrument might find more widespread application as a psychometric assessment tool that measures teachers' competence directly. In the light of recent developments in the area of teacher education, selection, and accountability—which have raised questions about the competences to be transmitted in teacher education, how schools or districts can evaluate the quality of their teachers, and how to provide teachers with feedback on their strengths and weaknesses—this aspect is of increasing importance. To date, most assessments of teacher quality rely on distal indicators such as university courses, degrees, or grades (Zumwalt & Craig, 2005). Our research identifies another way of gauging teacher qualifications in terms of the assets that seem most important for their primary task of teaching. Due to its pioneer character, our instrument is not yet suitable for use in high-stakes situations that require utmost reliability in identifying different levels of competence. We do not yet know enough about issues such as retest reliability or suitability for other samples, but addressing these questions is an important objective of our ongoing research agenda.

Second, our study provides some valuable insights into the "long arm" of university teacher training. Although our analyses suggest that the two teacher groups probably differed in certain background variables even prior to teacher training, they also indicate that there must be something specific about either GY teacher training or professional development at GY schools—over and above the different starting levels—that facilitates more extensive knowledge development in this context. Moreover, because no positive correlation was found between years of teaching practice and the two knowledge categories (see Brunner et al., 2006), teacher training can be assumed to be at the core of the development of the two knowledge categories. Future research may be able to provide deeper insights into the acquisition of PCK and CK during teacher training. For instance, longitudinal implementation of our tests at several critical stages in teacher education might provide more accurate information on the timing (e.g., in which phase of teacher education are PCK and CK acquired?) and mechanisms (e.g., which is needed to acquire the other?) of professional expertise development. Such studies may help to create instructional programs (at university and in the classroom) to foster the CK and PCK of student teachers, and to monitor their learning progress with respect to these knowledge categories.

The limitations of our study raise further interesting research questions. First, our study can only provide limited insights into the external validity of our measures. The finding that the teachers with more in-depth training in mathematics scored significantly higher in the content knowledge test may be seen as a first indication of the measure's external validity, but other approaches to the validity issue are also required. For instance, the convergent and discriminant validity of our measures should be investigated by employing our instrument in combination with other direct measures of mathematics teachers' professional knowledge (e.g., the newly developed standardized PRAXIS series; Educational Testing Service, 2006). Even more important, because our knowledge measures were assessed in a standardized testing situation, their implications for authentic learning situations remain to be investigated. Additional research is needed to examine precisely how PCK and CK regulate teaching behavior, and crucially their impact on student learning. Drawing on previous studies (e.g., Fennema & Franke, 1992; Ma, 1999), teachers with higher PCK might be expected to display a broader repertoire of instructional strategies and to be more likely to create cognitively stimulating learning situations. Indeed, first empirical findings indicate that teachers with higher PCK scores on our test tend to set tasks with higher potential for cognitive activation but do not seem to differ from their peers in terms of classroom management (Kunter et al., 2007). These findings provide first evidence for the discriminant validity of our instrument (for the approach of investigating the construct validty of PCK and CK by examining other populations, e.g., subject-matter experts or biology and chemistry teachers, with the tests, see Krauss, Baumert, & Blum, in press). Finally, strictly speaking, the generalizability of our results is limited to secondary mathematics teachers in Germany. Before final conclusions can be drawn about the dimensionality of teacher knowledge, our results need to be replicated in other samples; for instance, with teachers from countries with different educational systems. Last but not least, it is our hope that the present article might not only activate discussion on the professional knowledge of mathematics teachers, but also initiate similar endeavors for other school subjects.

References

Ball, D. L., Hill, H. C., & Bass, H. (2005, Fall). Knowing mathematics for teaching. *American Educator,* 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 ed., pp. 433–456). New York: Macmillan.

Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research, 35,* 463–482.

Brunner, M., Kunter, M., Krauss, S., Baumert, J., Blum, W., Dubberke, T., et al. (2006). Welche Zusammenhänge bestehen zwischen dem fachspezifischen Professionswissen von Mathematiklehrkräften und ihrer Ausbildung sowie beruflichen Fortbildung? [How is mathematics teachers' subject-specific professional knowledge related to their initial training and in-service training?] *Zeitschrift für Erziehungswissenschaft, 4,* 521–544.

Byrne, B. M., Shavelson, R. J., & Muthén, B. O. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105,* 456–466.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121–152.

Christiansen, B., & Walther, G. (1986). Task and activity. In B. Christiansen, A. G. Howson, & M. Otte (Eds.), *Perspectives on mathematics education* (pp. 243–307). Dordrecht, the Netherlands: Reidel.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159.

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

Educational Testing Service. (2006). *Proper use of the Praxis series and related assessments.* Retrieved June 19, 2006, from http://www.ets.org/Media/Tests/PRAXIS/pdf/guidelines.pdf

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

Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education.* New York: Teachers College Press.

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

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

Jackson, D. L. (2003). Revisiting sample size and number of parameter estimates: Some support for the N:q hypothesis. *Structural Equation Modeling, 10,* 128–141.

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

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.

Krauss, S., Baumert, J., & Blum, W. (in press). Secondary mathematics teachers' pedalogical content knowledge and content knowledge: Validation of the COACTIV constructs. *The International Journal on Mathematics Education.*

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. 39–59.) Münster, Germany: Waxmann.

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

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research, 32,* 53–76.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighting the merits. *Structural Equation Modeling, 9,* 151–173.

Ma, L. (1999). Knowing and Teaching Elementary Mathematics. Teachers' Understanding of Fundamental Mathematics in China and the United States. Mahwah, NJ: Erlbaum.

Marsh, H. W., Hau, K. T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33,* 181–220.

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.

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

Mayer, R. E. (2004b). Teaching of subject matter. *Annual Review of Psychology, 55,* 715–744.

Muthén, B. O. (1998–2004). *Mplus technical appendices.* Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (1998–2006). *Mplus user's guide* (4th ed.). Los Angeles, CA: Muthén & Muthén.

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62,* 307–332.

Palmer, D. J., Stough, L. M., Burdenski, T. K., Jr., & Gonzales, M. (2005). Identifying teacher expertise: An examination of researchers' decision making. *Educational Psychologist, 40,* 13–25.

Phelps, G., & Schilling, S. (2004). Developing measures of content knowledge for teaching reading. *Elementary School Journal, 105,* 31–48.

Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika, 50,* 83–90.

Schmidt, H. G., & Boshuizen, H. P. A. (1992). Encapsulation of biomedical knowledge. In D. A. Evans & V. L. Patel (Eds.), *Advanced models of cognition for medical training and practice* (pp. 265–282). New York: Springer.

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

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

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

Silver, E. A., Ghousseini, H., Gosen, D., Charalambous, C., & Strawhun, B. (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.

Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist, 61,* 394–403.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3,* 4–70.

Williams, G. (2002). Identifying tasks that promote creative thinking in mathematics: A tool. In B. Barton, K. Irwin, M. Pfannkuch, & M. Thomas (Eds.), *Mathematical education in the South Pacific: Proceedings of the 25th Annual Conference of the Mathematics Education Research Group of Australasia: Vol. II* (pp. 698–705). Auckland, New Zealand: MERGA.

Zumwalt, K., & Craig, E. (2005). Teachers' Characteristics: Research on the Indicators of Quality. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying Teacher Education: The Report of the AERA Panel on Research and Teacher Education* (pp. 157–260). Mahwah, NJ: Erlbaum.

Received August 29, 2006
Revision received October 4, 2007
Accepted October 7, 2007 ■