

# Teacher Education Effectiveness: Quality and Equity of Future Primary and Future Lower Secondary Teachers' General Pedagogical Knowledge

J. König, S. Blömeke, L. Paine, W.H. Schmidt, and F.-J. Hsieh

**Abstract** For more than two decades, three components of teacher knowledge have been discussed, namely, content knowledge (CK), pedagogical content knowledge (PCK), and general pedagogical knowledge (GPK). Although there is a growing body of analytic clarification and empirical testing with regard to CK and PCK, especially with a focus on mathematics teachers, hardly any attempt has been made to learn more about teachers' GPK. In the context of the Teacher Education and Development Study in Mathematics (TEDS-M), Germany, Taiwan, and the United States worked on closing this research gap by conceptualizing a theoretical framework and developing a standardized test of GPK, which was taken by representative samples of future elementary and middle school teachers in these countries. Four task-based subdimensions of GPK and three cognitive subdimensions of GPK were distinguished in this test. TEDS-M data are used (a) to test the hypothesis that GPK is not homogenous but multidimensional and (b) to compare the achievement of future elementary and middle school teachers in Germany, Taiwan and the US. The data revealed that US future teachers were outperformed by both the other groups. They showed a relative strength in one of the cognitive subdimensions, generating

---

Expanded version of König et al. (2011).

J. König (✉)

University of Cologne, Gronewaldstr. 2, 50931 Cologne, Germany

e-mail: [johannes.koenig@uni-koeln.de](mailto:johannes.koenig@uni-koeln.de)

S. Blömeke

Humboldt-University of Berlin, Berlin, Germany

L. Paine

Department of Teacher Education, College of Education, Michigan State University,  
East Lansing, MI 48824, USA

W.H. Schmidt

620 Farm Lane, Erickson Hall, Room 238, Michigan State University, East Lansing, MI 48824,  
USA

F.-J. Hsieh

National Taiwan Normal University, Taipei, Taiwan

strategies to perform in the classroom, indicating that in particular they had acquired procedural GPK during teacher education.

**Keywords** Adaptivity · Assessment · Classroom management · Cognitive process · General pedagogy · General pedagogical knowledge · GPK · Generating · Instruction · Motivation · Recall · Structure · Teaching methods

Researchers identify and distinguish among three domains of teacher knowledge (Baumert and Kunter 2006; Bromme 1997; Grossmann and Richert 1988; Shulman 1986, 1987): content knowledge (CK), pedagogical content knowledge (PCK), and general pedagogical knowledge (GPK). Regarding the latter domain, one can state that although over the past few years the body of research on teacher knowledge has been growing, it still remains an open question what exactly is meant by the term GPK and what this knowledge domain incorporates. In a time of globalization when the discourse on teacher education and the definition of what pre-service and in-service teachers have to know and be able to do are no longer limited to institutional, regional or national boundaries, the fact that the term itself is not used in all countries or at least not in the same way will inevitably come to the front and increasingly lead to the need for clarification.

Discussions about the reform of teacher education are often dominated more by normative than evidence-based statements (Ball et al. 2008; König and Blömeke 2013, *in press*). Especially with regard to general pedagogy as a component of teacher education programs, broad claims about its “uselessness” as well as about what future teachers need to know at the end of their training have been made and linked with requests either to eliminate this component or to structure it in a new way (Grossman 1992; Kagan 1992). Even if such discussions and assumptions may provide promising hypotheses, without empirical testing they have their limits in the process of improving teacher education (Larcher and Oelkers 2004).

The growing body of research in the field of (future) teachers’ knowledge has a special focus on subject-related issues—mostly exemplified by mathematics teachers in prominent research studies like *Learning Mathematics for Teaching* (LMT; e.g., Ball et al. 2008; Hill et al. 2004), *Mathematics Teaching in the 21st Century* (MT21; Schmidt et al. 2013) or *Professional Competence of Teachers, Cognitively Activating Instruction, and Development of Students’ Mathematical Literacy* (COACTIV; Baumert et al. 2010; Krauss et al. 2008). These studies mainly focused on CK and PCK in mathematics.

Also in the “*Teacher Education and Development Study: Learning to Teach Mathematics* (TEDS-M)”, the first comparative study on tertiary education with representative samples and direct testing of teacher knowledge, the common international questionnaire focused on the future teachers’ mathematics content knowledge (MCK) and mathematics pedagogical content knowledge (MPCK) in their last year of training. Three participating countries—the USA, Germany, and Taiwan—decided therefore to develop a national option measuring future teachers’ GPK. The

option was developed under the leadership of the German TEDS-M team (König and Blömeke 2009, 2010a, 2010b, 2010c; Blömeke and König 2010a, 2010b).

In this chapter, we report first how the general pedagogical knowledge test was conceptualized. It specifies the elements of GPK which future teachers have to acquire in teacher education in order to progress from the stage of teacher “novices” to “advanced beginners” (Berliner 2001, 2004). Based on data from future teachers in the USA, Germany, and Taiwan, the structure of GPK as well as specific strengths and weaknesses of future primary and future lower secondary teachers will be examined in a second step. Third, we will draw conclusions for next research steps and the reform of teacher education.

## 1 Defining General Pedagogical Knowledge of Future Teachers

Seen from an international perspective, it is a great challenge to determine what is meant by the term GPK and what this knowledge domain incorporates. In the USA, two broad labels—“educational foundations” and “teaching methods”—are needed to cover what may be labelled as “general pedagogy” in another country. In yet another country, the theoretical underpinnings of education may be provided by educational psychology, sociology of education or history of education. The opportunities to learn (OTL) implemented in these components of teacher education may be very diverse, too, not only across countries but also within one country.

The shape of general pedagogy is probably influenced by cultural perspectives on the objectives of schooling and on the role of teachers (Hopmann and Riquarts 1995). But at the same time evidence exists that there is some communality in the OTL due to the nature of teaching (Blömeke 2012; Blömeke and Kaiser 2012). A literature review reveals that two tasks of teachers are regarded as core tasks in almost all countries: instruction and classroom management. Generic theories and methods of instruction and learning as well as of classroom management can therefore be defined as essential parts of GPK. Less agreement exists as to what extent and what kind of knowledge about counselling or nurturing students’ social and moral development or knowledge about school management should also be included in the area of general pedagogy.

According to Shulman (1987, p. 8) general pedagogical knowledge involves “broad principles and strategies of classroom management and organization that appear to transcend subject matter” as well as knowledge about learners and learning, assessment, and educational contexts and purposes. Similarly, and extending this definition, Grossman and Richert (1988, p. 54) stated that GPK “includes knowledge of theories and learning and general principles of instruction, an understanding of the various philosophies of education, general knowledge about learners, and knowledge of the principles and techniques of classroom management.” Future teachers need to draw on this range of knowledge and weave it into coherent understandings and skills if they are to become competent to deal with what McDonald (1992) called the “wild triangle” that connects learner, subject matter and teacher in the classroom.

Since there was a lack of empirical studies on (future) teachers' GPK (Wilson and Berne 1999) when TEDS-M started, many key questions were at that time unanswered. There were virtually no studies showing how to fill these relatively broad domains of GPK so that one could develop items and actually test teachers (Baumert and Kunter 2006). Another open question was how to discriminate GPK from MPCK. In the USA, Germany and Switzerland, some first attempts existed to measure the GPK of future or practicing teachers (Baer et al. 2007; Grossman 1992; Schulte 2007) but these studies were restricted to specific institutions, languages or regions. Other studies had tried to capture GPK with self-reports of future teachers (e.g., Oser and Oelkers 2001) but these did not include objective tests.

Against the background of this research gap, the authors of this chapter aimed at developing a theoretical framework of future teachers' GPK that could be tested empirically across countries in the context of TEDS-M. Due to the complexity of GPK, the audience of an international survey, the target population of future teachers (and not practicing teachers) and with regard to standardized test procedures on a large scale, it was necessary to make certain restrictions in the definition of general pedagogy.

Following the concept of "competence" (see in general Weinert 2001; specified for the teaching profession by Bromme 1992, 1997, 2001), the study's framework focused on the mastering of professional tasks and reaching important objectives of teaching. This meant that the theoretical framework of GPK was structured in a task-based way and explicitly not according to the formal structure of general pedagogy as an academic discipline. Since it is widely accepted that instruction represents the core activity of teachers (Baumert and Kunter 2006; Berliner 2001, 2004; Blömeke et al. 2008; Bromme 1997), the central demands placed on teachers are related to student learning. While other teacher tasks like counselling or nurturing students' social and moral development were regarded equally important, they could not be included within the framework of TEDS-M. The cultural differences not only across the three participating countries but also within these did not suggest a common sense of "correct" or "incorrect" performance strategies necessary for objective testing. Clearly, these restrictions leave space for future research that follows a broader understanding of GPK.

## 2 Subareas of General Pedagogical Knowledge

Our focus on instruction served as a heuristic to select different topics and cognitive demands of general pedagogical knowledge. In order to operationalize it, we referred to the extensive research on instruction. Instructional research provides various models of school learning (e.g., Carroll 1963; Bloom 1976). Such models contain elements that are directly under the control of the teacher (e.g., the effectiveness with which a lesson is actually delivered) and elements that are characteristics of the students which are difficult to change (e.g., the students' general abilities to learn). Since our perspective focused on elements teachers can influence, we decided to

Content dimensions	Topics covered by the test items
structure	- structuring of learning objectives
	- lesson planning and structuring the lesson process
	- lesson evaluation
motivation/ classroom management	- achievement motivation
	- strategies to motivate single students/ the whole group
	- strategies to prevent and counteract interferences
adaptivity	- effective use of allocated time/ routines
	- strategies of differentiation
assessment	- use of a wide range of teaching methods
	- assessment types and functions
	- evaluation criteria
	- teacher expectation effects

**Fig. 1** Content dimensions and topics covered in the TEDS-M test of GPK

take the QAIT model by Slavin (1994) as a basis to describe teacher tasks in more detail.

The QAIT model is a model of effective instruction which focuses on four elements:

- The first element, “Quality of instruction” (Q), refers to activities of teaching that make sense to students, for instance, presenting information in an organized way or noting transitions to new topics.
- “Appropriate Levels of instruction” (A) is an element that refers to dealing with a heterogeneous class. For teachers, it is challenging to adapt instruction to students’ diverse needs. adaptivity deals, for instance, with the level of instruction (that is appropriate when a lesson is neither too difficult nor too easy for students) or with the different methods of within-class ability grouping.
- The third element, called “Incentives” (I), deals with the motivation of students to pay attention, to study, and to perform the tasks assigned. For a teacher, this means, for instance, relating topics to students’ experiences.
- “Time” (T) is the fourth element of the model. It refers to the quantitative aspect of instruction and learning, e.g., strategies of classroom management enabling students to spend a high amount of time on tasks.

According to Slavin (1994), the four elements are linked to each other and instruction is only then effective if they are all applied. The QAIT elements correspond to elements of other models and listings of effective teaching (Helmke 2003; Baumert et al. 2004; Good and Brophy 2007). So, the four elements can in fact be regarded as basic dimensions of teaching quality (Brophy 1999).

However, to identify potential shortcoming of this framework, we compared the basic dimensions of teaching quality with didactical points of view (cf. Klafki 1985; Tulodziecki et al. 2004; Good and Brophy 2007). In these, diagnosing and assessing student achievement was in fact more strongly focused. In the QAIT model it is more regarded a precondition of adaptivity. However, we specifically added it because assessing students is an essential teacher task (Good and Brophy 2007).

The approach of combining findings from instructional research and didactics led us to conceptualize GPK for teaching as is shown in Fig. 1. Teacher education was regarded as effective if future teachers in their last year of their training had acquired general pedagogical knowledge allowing them to prepare, structure, and evaluate lessons (“structure”), to motivate and support students as well as manage the classroom (“motivation/classroom management”), to deal with heterogeneous learning groups in the classroom (“adaptivity”), and to diagnose and assess student achievement (“assessment”). Three of these content dimensions corresponded to the QAIT model. Assessment was added due to its didactical relevance.

Apart from the task-based content dimensions of GPK, we defined dimensions of cognitive processes describing the cognitive demands on future teachers when they respond to test items. Following Anderson’s and Krathwohl’s elaborate and well-known model (Anderson and Krathwohl 2001), we distinguished three cognitive processes which summarized the original six processes: recalling, understanding/analyzing, and generating. Future teachers had to retrieve information from long-term memory in order to respond to a test item. They had to understand or to analyze a concept, a specific term or a phenomenon outlined by a specific test item. And they were asked to generate concrete strategies on how they would solve a typical classroom situation problem which includes evaluating this situation. Our hypothesis was that future teacher performance on test items varies according to these cognitive processes.

Distinguishing between declarative and procedural knowledge as another cognitive approach is very common in teacher research (besides Anderson and Krathwohl 2001 see e.g., Fenstermacher 1994; Bromme 2001). In our instrument, test items requiring future teachers to recall information predominantly measured declarative knowledge (“knowing that . . .”) including factual and conceptual knowledge while test items requiring future teachers to generate strategies not only measured declarative but also procedural knowledge (“knowing how . . .”). Procedural knowledge is of a situated nature (Putnam and Borko 2000).

### 3 Test Instrument

Content dimensions of GPK and cognitive demands made up a matrix which served as a heuristic for item development (see Fig. 2). For each cell, a subset of items was developed. Several expert reviews in the USA, Germany, and Taiwan as well as two large pilot studies were carried out. All experts who participated in the first item review which aimed at selecting items for the first pilot study testing a large pool of items were teacher educators in the field of general pedagogy. Their research had to be related to the topic of teacher knowledge and they had to be at least PhD candidates. Experts that participated in the second and following reviews which aimed at selecting items for the final test instrument according to specific criteria or that aimed at validating the test instrument, respectively, had to endow a university chair with a specialization on research about teacher knowledge. Based on these review

	recall	Understand/ analyze	Generate
structure			
motivation/classroom management			
adaptivity			
assessment			

**Fig. 2** Test design matrix

processes and empirical findings from the two pilot studies (e.g., item parameter estimates) as well as on conceptual considerations with respect to the framework, the final item set was selected (König and Blömeke 2009, 2010a, 2010c; Blömeke and König 2010a).

The TEDS-M test measuring GPK of future primary teachers consisted of 85 test items. The TEDS-M test measuring GPK of future lower secondary teachers consisted of 77 test items. In both cases, we used dichotomous and partial-credit items, open-response (about half of the items) and multiple-choice items. Items were fairly equally distributed across the four content dimensions and the three cognitive dimensions.

Following the TEDS-M test design for MCK and MPCK (for details see Tatto et al. 2008), a balanced incomplete block-design (Adams and Wu 2002; von Davier et al. 2006) with five booklets for the future primary teacher survey and three booklets for the future secondary teacher survey was used so that each person had to respond to only 60–65 % of all test items. With permission of the TEDS-M International Study Center, the GPK test was added at the end of the original TEDS-M future teacher questionnaire.

Germany and the USA used an identical booklet design in their primary and lower secondary survey instruments allowing future teachers 30 minutes to respond to the GPK test items. In contrast, Taiwan selected five complex test items due to limited survey time, covering each cell of our test design matrix and corresponding to criteria such as difficulty level, estimated response time and item discrimination. In addition, Taiwan decided to implement the GPK in the lower secondary future teacher survey only.

Item-Response-Theory (IRT) scaling methods were used to estimate scores across the different booklets for the primary and the lower secondary future teacher survey, respectively. With the methods implemented in a software package like *Conquest* (Wu et al. 1997), it is possible to create reliable achievement scores even if a person has only responded to a selection of test items if this selection was done rigorously according to a range of specific criteria. Technical issues of the test instrument such as the booklet design were reviewed by experts from each country as well. These had to have at least a PhD in psychometrics, most of them had university chairs.

Which of the following cases represents an example of intrinsic motivation, and which represents an example of extrinsic motivation?		
	<i>Check <u>one</u> box in each row.</i>	
	<b>intrinsic motivation</b>	<b>extrinsic motivation</b>
<b>A student learns before a test in mathematics, because he/she ...</b>		
A.	expects a reward for a good grade.	<input type="checkbox"/> <sub>1</sub> <input type="checkbox"/> <sub>2</sub>
B.	wants to avoid the consequences of a bad grade.	<input type="checkbox"/> <sub>1</sub> <input type="checkbox"/> <sub>2</sub>
C.	is interested in problems of mathematics.	<input type="checkbox"/> <sub>1</sub> <input type="checkbox"/> <sub>2</sub>
D.	does not want to disappoint his/her parents.	<input type="checkbox"/> <sub>1</sub> <input type="checkbox"/> <sub>2</sub>
E.	wants to maintain his/her relative rank in the class.	<input type="checkbox"/> <sub>1</sub> <input type="checkbox"/> <sub>2</sub>

**Fig. 3** Item example for GPK about “motivation” and “analyze”

Imagine you are helping a future teacher to evaluate her lesson because she has never done this before. To help her adequately analyze her lesson, what question would you ask? Formulate ten essential questions and write them down.

**Fig. 4** Item example for GPK about “structure” and “generate”

Two item examples (see Figs. 3 and 4) illustrate the GPK test.<sup>1</sup> The first item measured knowledge about “motivating” students. Future teachers had to recall basic terminology of achievement motivation (“intrinsic motivation” and “extrinsic motivation”) and they were asked to analyze five statements against the background of this distinction. Statement C represented an example of “intrinsic motivation” whereas A, B, D, and E were examples for “extrinsic motivation”.

The second item example (see Fig. 4) was an open-response item. Here, future teachers were asked to support another future teacher and evaluate her lesson. This is a typical challenge during a peer-led teacher education practicum, but practicing teachers are also regularly required to analyze and reflect on their own as well as their colleagues’ lessons. The item measured knowledge of “structuring” lessons. The predominant cognitive process was to “generate” fruitful questions.

For the open-response items, coding rubrics were developed and reviewed by experts on teacher education in the USA, Germany, and Taiwan to avoid culturally biased coding and scoring. The coding instructions were developed in an extensive interplay of deductive (from our theoretical framework) and inductive approaches (from empirical teacher responses). In a pilot phase, codes from several independent

---

<sup>1</sup>Since we plan to use the test instrument in future studies, we cannot include more item examples in this article. If other researchers are interested in doing research in this field, we will be pleased to provide a complete documentation of a shorter version of the instrument with half of the test item pool, including various materials such as coding rubrics, scoring instruction, empirically based information on item parameters, and test booklets. This documentation (König and Blömeke 2010c) allows to use the GPK test independently from the authors who developed it. Please contact the initial author of this article for further information.



- |   |
|---|
| <p>1) <i>Do your students have prior knowledge about the subject?</i></p> <p>2) <i>What are your objectives?</i></p> <p>3) <i>Are the students working individually or in groups?</i></p> <p>...</p> <p>10) <i>Have your students gained the knowledge from the lesson?</i></p> |
|---|

**Fig. 5** A US future teacher’s response to the item presented in Fig. 4

raters were discussed and coding instructions were revised and expanded. The result was then reviewed by experts. Thus, the coding manual is theoretically based as well as data-based. The codes were intended to be low-inferent, i.e. every response was coded with the least possible amount of inferences by the raters.

All questionnaires were coded by two raters independently of each other on the basis of the coding manual. As a measure of consensus and internal consistency, Cohen’s Kappa was estimated (Jonsson and Svingby 2007). It ranges from 0.80 to 0.99 with an average of  $M = 0.91$  ( $SD = 0.07$ ). This can be regarded as a good result. If the raters did not agree, agreement was achieved in joint discussions, calling on a third rater if necessary.

After having established reliable and culturally unbiased coding schemes, a scoring strategy for complex open-response items was developed to decide which codes should be rewarded and which could not because they were not appropriate. Again, experts from the three countries had to agree which codes would appropriately reflect outcomes of their teacher education systems. Illustrating this strategy with the test item shown in Fig. 4, codes were scored as appropriate if they addressed the “context” of the lesson (e.g., prior knowledge of students), the “input” (e.g., objectives of the lesson), the “process” (e.g., teaching methods used), or the “output” of the lesson (e.g. student achievement). The original answer given by a future teacher from the USA is an example for the scoring strategy (see Fig. 5).

## 4 Research Questions and Data Analyses

Two main research questions lead the present article: How is the general pedagogical knowledge of future primary and lower secondary teachers structured? Which level of achievement did future teachers from the USA, Germany and Taiwan show?

Our theoretical framework (see Fig. 2) had outlined four content and three cognitive dimensions of GPK. This means we assumed that GPK is multidimensional. An alternative hypothesis would be that GPK is homogeneous or one-dimensional. Technically spoken, the latter would imply an IRT scaling model in which only one latent variable was specified by all test items. Model 1 in Fig. 6 shows a graphical representation of this idea. Certain psychometric indicators (which are described below in detail) can be used as criteria to evaluate the alternative models.

Multi-dimensional IRT scaling models can be applied to models 2 and 3 in Fig. 6 which hypothesize that it is possible to distinguish four content dimensions

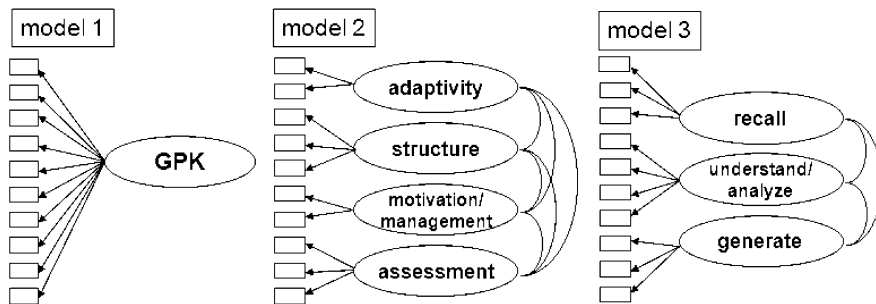


Fig. 6 One-dimensional and multi-dimensional modelling of GPK

(structure, motivation/management, adaptivity, assessment) and three cognitive dimensions (recalling, understanding/analyzing, generating). We specified a four-dimensional model with four latent variables and a three-dimensional model with three latent variables. All analyses were done with the IRT software *Conquest* (Wu et al. 1997; Wu 1997).

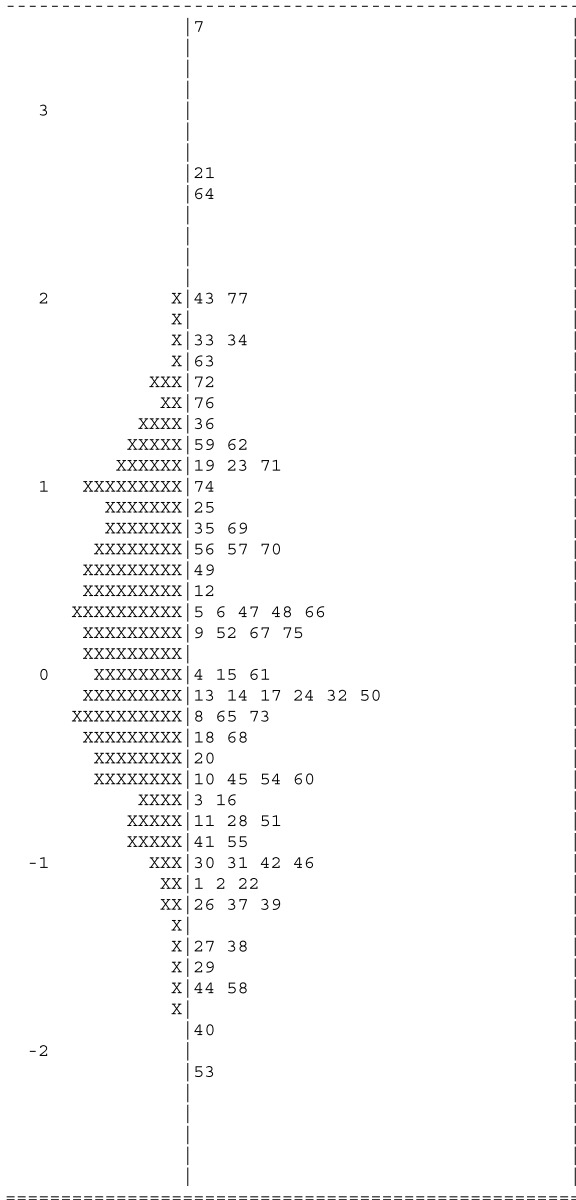
Findings on this research question provide information as to how we should report GPK results when the achievement future teachers from different countries is compared: as one overall score or as separate subscores. If reliable sub-dimensions of GPK can be modelled, we would be able to compare the outcomes of teacher education in the USA, Germany, and Taiwan in more detail because strengths and weaknesses of GPK could be described along content domains and cognitive processes. Findings of this kind would provide insight into the effectiveness of teacher education systems across countries. In a time of globalization and international competition, this is becoming increasingly important when discussing reforms of teacher education.

## 5 Empirical Findings on the Structure of GPK

Figure 7 shows an item-person map from the uni-dimensional IRT analysis (future lower secondary teacher survey). On the left side, abilities of future teachers are represented (one “X” represents eight persons), while on the right side the distribution of test items is shown (each of the 77 test items has a number). If the location of an item and a person match, the person has a probability of 0.5 to succeed on that item. The higher a person is above an item on the scale, the more likely the person will succeed on the item. The lower a person is below an item on the scale, the more likely the person will be unsuccessful on the item.

The map reveals that the GPK test covered the TEDS-M sample of future lower secondary teachers from the USA, Germany and Taiwan quite well as the range of person abilities (left side) was well covered by item difficulties (right side). The one-

**Fig. 7** Item-person map of one-dimensional Rasch-scaling



dimensional model and its results showed that it was possible to create an overall GPK test score. The reliability was good (EAP reliability 0.78).<sup>2</sup>

<sup>2</sup>Another psychometric indicator is the item fit statistics. The weighted mean squares mainly ranged from 0.80 to 1.20 with a few exceptions, which is a good result (Adams 2002; Wright

Model 2	Future primary teachers			Future lower secondary teachers		
	(1)	(2)	(3)	(1)	(2)	(3)
(1) adaptivity						
(2) structure	.72			.60		
(3) motivation/management	.77	.77		.63	.57	
(4) assessment	.81	.74	.81	.69	.65	.80

**Fig. 8** Intercorrelations of GPK content domains

In contrast to a model in which all items measure one latent ability (see model 1 in Fig. 7), in a multi-dimensional IRT model test items were scaled as documented in our conceptual framework according to their content (structure, motivation/classroom management, adaptivity, assessment; see model 2 in Fig. 6) or, alternatively, according to the cognitive processes requested (recalling, understanding/analyzing, generating; see model 3 in Fig. 6).<sup>3</sup> The reliability estimates of the content domains were lower than the reliability of the overall GPK score but mostly acceptable (future primary teachers: 0.79 for structure, 0.78 for adaptivity, 0.74 for motivation/classroom management, 0.72 for assessment; future lower secondary teachers: 0.70, 0.72, 0.65, 0.64).

The intercorrelations of the four domains were high but they did not indicate homogeneity taking into account that they did not include measurement error (see Fig. 8). This result represents an indicator to assume the hypothesized multidimensionality rather than uni-dimensionality of future teachers' GPK.<sup>4</sup> Interestingly, "assessment" and "motivation/classroom management" showed the highest intercorrelation. This pattern might mirror coherence of corresponding opportunities to learn in teacher education in contrast to "structure" and "adaptivity" which seems to be more distant to the area of assessment.

The cognitive dimensions were also modelled as hypothesized. The reliability of each of these three subscales was acceptable for "recalling" and "understanding/analyzing". The reliability of "generating" was rather low, however, indicating that this dimension was difficult to measure. Again, intercorrelations between the different cognitive domains were partly relatively low showing that GPK is a heterogeneous construct (see Fig. 9).

Recalling and understanding/analyzing seem to be well interconnected cognitive processes (almost 0.80). By contrast, they both were only loosely connected with the third cognitive demand of generating (less than 0.50). Although its lower reliability

et al. 1994). Exceptions occurred when an item could not be excluded for theoretical reasons. Then we accepted weighted mean squares ranging from 0.75 to 0.79 and from 1.21 to 1.25.

<sup>3</sup>To investigate the internal consistency of the four dimensions we used "Expected a posteriori estimation (EAP)" as these parameters deliver an unbiased estimation of the population and take into account the multidimensional structure of the model (Wu 1997).

<sup>4</sup>The item fit indices generally showed a slight improvement in the four-dimensional model compared to the one-dimensional model. The deviance of the two models revealed that the four-dimensional model fits the data significantly better than the one-dimensional model. Thus, there were in fact several indications suggesting the hypothesized multidimensionality of future middle school teachers' GPK.

Model 3	Future primary teachers		Future lower secondary teachers	
	(1)	(2)	(1)	(2)
(1) recall				
(2) Understand/analyze	.76		.79	
(3) Generate	.69	.83	.49	.46

**Fig. 9** Intercorrelations of cognitive domains of GPK

Country	Future primary teachers			Future lower secondary teachers		
	M	SE	SD	M	SE	SD
Germany	601	3.7	95	576	4.9	85
Taiwan	-	-	-	572	3.2	52
International	500	0.7	100	500	2.2	100
USA	462	2.7	72	440	3.0	66

**Fig. 10** Overall GPK test score

has to be taken into account, we can hypothesize that GPK seems to consist of the ability to generate strategies as a response to typical classroom situation vignettes on the one hand and of declarative knowledge measured by items labelled as “recalling” or “understanding/analyzing” on the other hand.

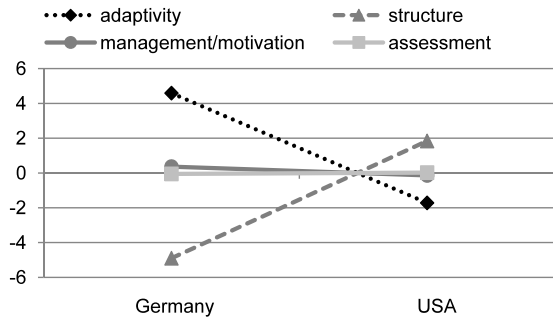
## 6 International Comparison of Future Teachers’ GPK

Our effort to measure GPK as an element of the professional knowledge of future teachers aimed at an international comparison. First, we present results on the overall GPK test score. To facilitate the reading, the mean was transformed to 500 test points with a standard deviation of 100 test points for each the primary and the lower secondary survey. Figure 10 shows the means, standard errors of the means, and the standard deviation for each country.

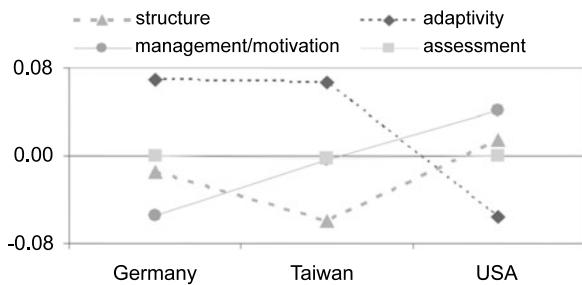
The data revealed that future teachers from Germany and Taiwan significantly outperformed their counterparts from the USA. The achievement of US future teachers was more than one and a half standard deviation lower than the achievement of German and, in the case of the future lower secondary teacher survey, Taiwanese future teachers. This is a difference which is of high practical relevance. There was no statistically significant difference between teacher achievement in Germany and Taiwan.

The next step in describing the GPK of future teachers is related to the subdimensions of this knowledge area. Because of the very large country mean differences on the overall GPK test score, we used ipsative measures to depict strengths and weaknesses of each country’s performance on the subdomains. Ipsative measures describe the *relative* achievement of future teachers in subdimensions; they

**Fig. 11** Relative strengths and weaknesses for content domains of GPK (future primary teachers)



**Fig. 12** Relative strengths and weaknesses for content domains of GPK (future lower secondary teachers)

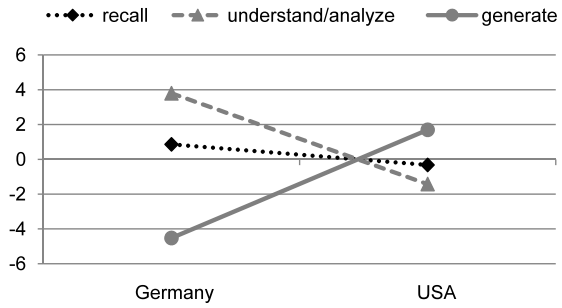


are standardized differences of each subdimension compared to the overall mean (the unit is one standard deviation).<sup>5</sup>

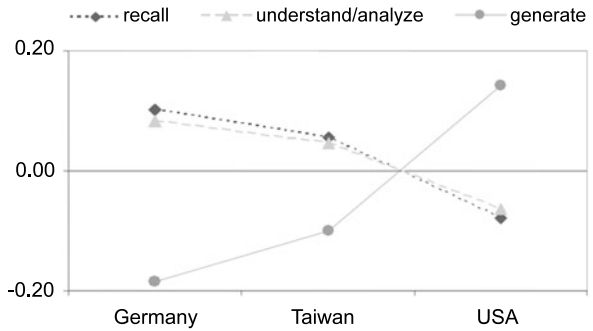
With regard to future lower secondary teachers, Fig. 12 reveals that US future teachers showed relatively equal performances on the three content domains structure, classroom management/motivation, and assessment (these scores were not significantly different from zero) whereas they showed a relatively weak performance on adaptivity. In contrast, future teachers in Taiwan and Germany showed a relatively high performance on this subdimension, leading to the assumption that they acquired general pedagogical knowledge to a particularly large extent about the use of a wide range of teaching methods in order to deal with heterogeneity in classroom situations. For the future primary teachers, adaptivity is in Germany similarly mirrored as strength (Fig. 11).

<sup>5</sup>Fischer (2004) once explained ipsative measures by using the following analogy: “Let us consider the example of a mouse and an elephant. Assume someone measured the extremities of both animals and used within-subject (within-animal) standardization [i.e. ipsative measures]. If the researcher would now proceed to compare the length of, let us say, the legs, probably no significant differences would be found. This is despite the fact that the legs of an elephant and a mouse are surely different. This is because all the measures are related to the size of the whole animal. [...] If we compare the tail of the mouse and the elephant using ipsative measures, we would probably conclude that the mouse’s tail is significantly longer than the tail of the elephant. It is important to note that this comparison makes sense only if we consider the length of the tail relative to the overall size of both animals. Obviously, relative to the overall size of the mouse and the elephant, the mouse’s tail is longer than the tail of the elephant.”

**Fig. 13** Relative strengths and weaknesses for cognitive processes (future primary teachers)



**Fig. 14** Relative strengths and weaknesses for cognitive processes (future lower secondary teachers)



Figures 13 and 14 depict the relative country differences with respect to cognitive processes. Future teachers in Germany showed a relatively weak performance in generating strategies. Compared with their performances in recalling GPK, understanding or analyzing a concept of general pedagogy related to teaching, they seemed to struggle when asked to evaluate classroom situations and to create adequate and multiple strategies to solve typical problems.

Taiwanese future teachers showed a balanced performance on these three cognitive processes (i.e. there were no statistically significant differences from zero). In contrast to Germany, US future teachers showed a relative strength in generating strategies related to typical classroom situations. Although their overall GPK performance was low compared with the performance of future teachers in Taiwan and Germany, the US GPK profile pointed to a suitable strength which would be very helpful in class. So, their declarative general pedagogical knowledge was below expectations whereas their procedural knowledge was above—however, the low mean has still to be taken into account.

## 7 Summary and Conclusions

General pedagogical knowledge (GPK) is a central component of teacher knowledge. Teacher education programs in many countries therefore provide corresponding opportunities to learn although they are sometimes labelled differently (“edu-

cational foundations”, “teaching methods”, “general pedagogy”, “educational psychology” etc.). In this paper we outlined a way to define and conceptualize general pedagogical knowledge for teaching in order to develop a test for large-scale assessments and international comparisons. Researchers from three countries—the USA, Germany and Taiwan—worked together in order to achieve validity with respect to teacher tasks and to avoid cultural bias.

According to our findings and in contrast to our hypothesis, it is legitimate to regard GPK as a homogenous construct. However, several indicators revealed that it is at the same time appropriate to distinguish between four content domains (structure, adaptivity, classroom management/motivation, assessment) and three cognitive processes (recalling, understanding/analyzing, generating) as we had assumed when we had developed our theoretical framework. Technically speaking, multi-dimensional IRT models fit the data significantly better than a one-dimensional model. Conceptually speaking, knowledge in one of these subdimensions does not necessarily mean an equal amount of knowledge in another subdimension. Since using the multi-dimensional estimates, future teachers’ performance can be described in more detail, we reported both approaches.

US future teachers were significantly outperformed by future teacher in Germany and Taiwan with regard to the overall GPK test score. The difference of more than 1.5 standard deviations was very large. It meant that there was almost no overlap between US teachers on the one side and Germany and Taiwanese teachers on the other side. Most of the worst achieving teachers from the latter two groups did still better than most of the best achieving teachers from the USA. Neglecting this large mean difference, country-specific profiles revealed that US future teachers had a relative GPK strength in generating classroom strategies but a weakness in recalling knowledge and analyzing problems. Future teachers from Germany showed a contrary profile whereas that from Taiwan was balanced.

Such results provide important information about teacher education in the USA compared to teacher education in Germany and Taiwan. The data indicated that there were probably more opportunities in the latter two countries to acquire systematic (declarative) knowledge whereas the focus in US teacher education seemed to be on acquiring teaching skills and, by this mean, developing procedural knowledge. One may conclude that even if teaching skills is what finally matters in the classroom, there is some evidence that procedural knowledge and skills should be built on extensive and systematic factual and conceptual knowledge. The low mean achievement of US future teachers may be caused by their limits in systematic knowledge. Thus, if it were possible to increase the mean GPK level of US teachers but to keep their strength in generating classroom-based solutions, the advantages of both approaches would be realized. Such a reform would require a careful look at the curriculum of teacher education in the USA compared to other countries.

Our results revealed that a comparative approach enables us to move beyond the familiar, and to see teacher education of different countries with a kind of “peripheral vision” (Bateson 1994). Strengths and weaknesses become visible. The structure and the content of teacher education may depend on a deeper rationale which may be a result of cultural boundaries (see specifically with respect to teacher education in the US and Germany Blömeke and Paine 2008). Stigler and Hiebert (1999)



argue that teaching reflects “cultural scripts”. And like the fish who is not aware of the water in which it swims, cultural givens are too often invisible to our consideration as we debate research designs, in this case, for research about teacher education. As a consequence, there is an increasing demand for comparative information about teacher education programs (König and Blömeke 2013, [in press](#)).

However, from a methodological point of view comparative studies are challenging—not least in an ill-defined area like general pedagogy. Cross-country validity is a request hard to achieve. Evaluation designs and measurement instruments have to be developed, for example, in order to precisely examine the outcomes of teacher education on the basis of educational standards. A central deficit of teacher education research has for a long time been a too narrow focus on single institutions or even single classes (Cochran-Smith and Fries 2005; Risko et al. 2008). Broader perspectives provide more orientation.

Grossman and McDonald (2008), for example, make a persuasive argument for the need for teacher education research to move beyond its “adolescence” (185) by identifying common factors that allow shared and more precise language. They suggest the “progress . . . will require researchers . . . to reach outside their immediate communities, to look over their backyards to see and learn from what their neighbors are doing” (199). A measurement instrument that is valid across three culturally very different countries as we have presented in this chapter is therefore a very important step on the way to encounter this research deficit in a time of globalization.

## References

- Adams, R. (2002). *Scaling PISA Cognitive Data*. Paris: OECD.
- Adams, R., & Wu, M. (Eds.) (2002). PISA 2000 Technical Report. OECD. <http://www.oecd.org/dataoecd/53/19/33688233.pdf>. Accessed 1 March 2008.
- Anderson, L. W., & Krathwohl, D. R. (Eds.) (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Baer, M., Dörr, G., Fraefel, U., Kocher, M., Küster, O., Larcher, S., Müller, P., Sempert, W., & Wyss, C. (2007). Werden angehende Lehrpersonen durch das Studium kompetenter? *Unterrichtswissenschaft*, 35(1), 15–47.
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching. What makes it special? *Journal of Teacher Education*, 59(5), 389–407.
- Bateson, M. C. (1994). *Peripheral visions: Learning along the way*. New York: Harper Collins.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- Baumert, J., Kunter, M., Brunner, M., Krauss, S., Blum, W., & Neubrand, M. (2004). Mathematikunterricht aus Sicht der PISA-Schülerinnen und -Schüler und ihrer Lehrkräfte. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost, & U. Schiefele (Eds.), *Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs*, PISA, 2003 (pp. 314–354). Münster: Waxmann
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, Th., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1), 133–180.
- Berliner, D. C. (2001). Learning about and learning from expert teachers. *Educational Research*, 35, 463–482.

- Berliner, D. C. (2004). Describing the behavior and documenting the accomplishments of expert teachers. *Bulletin of Science, Technology & Society*, 24(3), 200–212.
- Blömeke, S. (2012). Content, professional preparation and teaching methods: how diverse is teacher education across countries? *Comparative Education Review*, 56(4), 684–714.
- Blömeke, S., & Kaiser, G. (2012). Homogeneity or heterogeneity? Profiles of opportunities to learn in primary teacher education and their relationship to cultural context and outcomes. *ZDM—The International Journal on Mathematics Education*, 44(3), 249–264.
- Blömeke, S., & König, J. (2010a). Messung des pädagogischen Wissens: Theoretischer Rahmen und Teststruktur. In S. Blömeke, G. Kaiser, & R. Lehmann (Eds.), *TEDS-M 2008 – Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte im internationalen Vergleich* (pp. 239–269). Münster: Waxmann
- Blömeke, S., & König, J. (2010b). Pädagogisches Wissen angehender Mathematiklehrkräfte im internationalen Vergleich. In S. Blömeke, G. Kaiser, & R. Lehmann (Eds.), *TEDS-M 2008 – Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte im internationalen Vergleich* (pp. 270–283). Münster: Waxmann
- Blömeke, S., & Paine, L. (2008). Getting the fish out of the water: considering benefits and problems of doing research on teacher education at an international level. *Teaching and Teacher Education*, 24(4), 2027–2037.
- Blömeke, S., Paine, L., Houang, R. T., Hsieh, F.-J., Schmidt, W. H., Tatto, M. T., Bankov, K., Cedillo, T., Cogan, L., Han, Sh.I., Santillan, M. N., & Schulle, J. (2008). Future teachers' competence to plan a lesson. First results of a six-country study on the efficiency of teacher education. *ZDM—The International Journal on Mathematics Education*, 40(5), 749–762.
- Bloom, B. S. (1976). *Human Characteristics and School Learning*. New York: McGraw-Hill.
- Bromme, R. (1992). *Der Lehrer als Experte: zur Psychologie des professionellen Wissens*. Bern: Hogrefe & Huber.
- Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In F. E. Weinert (Ed.), *Enzyklopädie der Psychologie: Psychologie des Unterrichts und der Schule* (vol. 3, pp. 177–212). Göttingen: Hogrefe & Huber
- Bromme, R. (2001). Teacher Expertise. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 15459–15465). Amsterdam: Elsevier.
- Brophy, J. (1999). Teaching. Brussels: International Academy of Education. <http://www.ibe.unesco.org/publications/EducationalPracticesSeriesPdf/prac01e.pdf>. Accessed 1 January 2008.
- Carroll, J. B. (1963). A model of school learning. *Teachers College Record*, 64, 723–733.
- Cochran-Smith, M., & Fries, K. (2005). Researching teacher education in changing times: politics and paradigms. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying Teacher Education: The Report of the AERA Panel on Research and Teacher Education* (pp. 69–110). Mahwah: Erlbaum.
- von Davier, A. A., Carstensen, C. H., & von Davier, M. (2006). Linking Competencies in Horizontal, Vertical, and Longitudinal Settings and Measuring Growth. In J. Hartig, E. Klieme, & D. Leutner (Eds.), *Assessment of Competencies in Educational Contexts* (pp. 53–80). Göttingen: Hogrefe & Huber
- Fenstermacher, G. D. (1994). The Knower and the Known: The Nature of Knowledge in Research on Teaching. In L. Darling-Hammond (Ed.), *Review of Research in Education* (vol. 20, pp. 3–56). Washington
- Fischer, R. (2004). Standardization to account for cross-cultural response bias: a classification of score adjustment procedures and review of research in JCCP. *Journal of Cross-Cultural Psychology*, 35(3), 263–282.
- Good, T. L., & Brophy, J. E. (2007). *Looking in Classrooms*, Boston: Allyn & Bacon.
- Grossman, P. L. (1992). Why models matter: an alternate view on professional growth in teaching. *Review of Educational Research*, 62(2), 171–179.
- Grossman, P., & McDonald, M. (2008). Back to the future: directions for research in teaching and teacher education. *American Educational Research Journal*, 45(1), 184–205.
- Grossman, P. L., & Richert, A. E. (1988). Unacknowledged knowledge growth: a re-examination of the effects of teacher education. *Teaching and Teacher Education*, 4(1), 53–62.

- Helmke, A. (2003). *Unterrichtsqualität erfassen, bewerten, verbessern*. Seelze: Kallmeyer.
- Hill, H. C., Rowan, B., & Loewenberg Ball, D. (2004). *Effects of Teacher's Mathematical Knowledge for Teaching on Student Achievement*. Paper presented at the American Educational Research Association, San Diego, CA.
- Hopmann, S. & Riquarts, K. (Eds.) (1995). *Didaktik und/oder Curriculum: Grundprobleme einer international-vergleichenden Didaktik*. Weinheim: Beltz.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130–144.
- Kagan, D. M. (1992). Professional growth among preservice and beginning teachers. *Review of Educational Research*, 62(2), 129–169.
- Klafki, W. (1985). *Neue Studien zur Bildungstheorie und Didaktik. Beiträge zur kritisch-konstruktiven Didaktik*. Weinheim: Beltz.
- König, J., & Blömeke, S. (2009). Pädagogisches Wissen von angehenden Lehrkräften: Erfassung und Struktur von Ergebnissen der fachübergreifenden Lehrerausbildung. *Zeitschrift für Erziehungswissenschaft*, 12(3), 499–527.
- König, J., & Blömeke, S. (2010a). Messung des pädagogischen Wissens: Theoretischer Rahmen und Teststruktur. In S. Blömeke, G. Kaiser, & R. Lehmann (Eds.), *TEDS-M 2008 – Professionelle Kompetenz und Lerngelegenheiten angehender Primarstufenlehrkräfte im internationalen Vergleich* (pp. 253–273). Münster: Waxmann
- König, J., & Blömeke, S. (2010b). Pädagogisches Wissen angehender Primarstufenlehrkräfte im internationalen Vergleich. In S. Blömeke, G. Kaiser, & R. Lehmann (Eds.), *TEDS-M 2008 – Professionelle Kompetenz und Lerngelegenheiten angehender Primarstufenlehrkräfte im internationalen Vergleich* (pp. 275–296). Münster: Waxmann
- König, J., & Blömeke, S. (2010c). Pädagogisches Unterrichtswissen (PUW). Dokumentation der Kurzfassung des TEDS-M-Testinstruments zur Kompetenzmessung in der ersten Phase der Lehrerausbildung. Berlin: Humboldt-University
- König, J., Blömeke, S., Paine, L., Schmidt, W. H., & Hsieh, F.-J. (2011). General pedagogical knowledge of future middle school teachers: on the complex ecology of teacher education in the United States, Germany, and Taiwan. *Journal of Teacher Education*, 62(2), 188–201.
- König, J., & Blömeke, S. (2013, in press). TEDS-M Country Report on Teacher Education in Germany. In M. T. Tatto (Ed.), *Policy, practice, and readiness to teach primary and secondary mathematics. The teacher education and development study in mathematics international report. Vol. 5: Encyclopedia*.
- Krauss, S., Baumert, J., & Blum, W. (2008). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: validation of the COACTIV constructs. *Zentralblatt für Didaktik der Mathematik*, 40, 873–892.
- Larcher, S., & Oelkers, J. (2004). Deutsche Lehrerbildung im internationalen Vergleich. In S. Blömeke, P. Reinhold, G. Tulodziecki, & J. Wildt (Eds.), *Handbuch Lehrerausbildung* (pp. 128–150). Bad Heilbrunn: Klinkhardt
- McDonald, J. P. (1992). *Teaching: Making sense of an uncertain craft*. New York: Teachers College Press.
- Oser, F., & Oelkers, J. (Eds.) (2001). *Die Wirksamkeit der Lehrerbildungssysteme*. Chur: Ruediger.
- Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning. *Educational Researcher*, 29, 4–15.
- Risko, V. J., Roller, C. M., Cummins, C., Bean, R. M., Block, C. C., Anders, P. L., & Flood, J. (2008). A critical analysis of research on reading teacher education. *Reading Research Quarterly*, 43(3), 252–288.
- Schmidt, W. H., Blömeke, S., & Tatto, M. T. (2013, in press). *Teacher preparation from an international perspective*. New York: Teachers College Press.
- Schulte, K. (2007). Selbstwirksamkeitserwartung und Professionswissen in der Lehramtsausbildung. Göttingen. [http://www.psych.uni-goettingen.de/special/grk1195/projekte/g/poster\\_projekt\\_g.pdf](http://www.psych.uni-goettingen.de/special/grk1195/projekte/g/poster_projekt_g.pdf). Accessed 29 January 2008.
- 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 Research*, 57, 1–22.
- Slavin, R. E. (1994). Quality, appropriateness, incentive, and time: a model of instructional effectiveness. *International Journal of Educational Research*, 21, 141–157.
- Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: Free Press.
- Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Peck, R., & Rowley, G. (2008). *Teacher Education and Development Study in Mathematics (TEDS-M): Conceptual framework*. East Lansing.
- Tulodziecki, G., Herzig, B., & Blömeke, S. (2004). *Gestaltung von Unterricht. Eine Einführung in die Didaktik*. Bad Heilbrunn: Klinkhardt.
- Weinert, F. E. (2001). A concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and Selecting Key Competencies* (pp. 45–65). Göttingen: Hogrefe & Huber
- Wilson, S. M., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of research in education* (pp. 173–210). Washington, DC
- Wright, B. D., Linacre, M., Gustafsson, J.-E., & Martin-Loff, P. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8, 370.
- Wu, M. L. (1997). *The development and application of a fit test for use with generalised item response model*. Masters of Education Dissertation. Melbourne.
- Wu, M. L., Adams, R. J., & Wilson, M. R. (1997). ConQuest: Multi-Aspect Test Software [computer program]. Camberwell, Vic.: Australian Council for Educational Research.