

DIRECTORATE FOR EDUCATION AND SKILLS
CENTRE FOR EDUCATIONAL RESEARCH AND INNOVATION (CERI) GOVERNING BOARD

INNOVATIVE TEACHING FOR EFFECTIVE LEARNING

Background Document: Designing an International Instrument to Assess Teachers' General Pedagogical Knowledge (GPK): Review of Studies, Considerations, and Recommendations

This paper was developed as part of the background research for the ITEL Teacher Knowledge Survey. The purpose of the paper is to review previous research and to make suggestions and recommendations for developing an international instrument for profiling teachers' general pedagogical knowledge. The work was commissioned to Prof. Dr. Johannes König of the University of Cologne, Germany. This is a revised version of the paper presented at the CERI Governing Board in April 2014.

The opinions expressed in this paper are the sole responsibility of the author and do not necessarily reflect those of the OECD or of the governments of its member countries.

Sonia Guerriero, Project Leader (email: sonia.guerriero@oecd.org)

JT03385200

Complete document available on OLIS in its original format

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

TABLE OF CONTENTS

SUMMARY4

INTRODUCTION.....5

 What this Report is Aiming at5

PART 1: BACKGROUND REVIEW AND POLICY CONTEXT6

 On the State of Research and the Justification for an Assessment of Teacher Knowledge6

PART 2: A REVIEW OF SOME PREVIOUSLY CONDUCTED STUDIES9

 Methodological Approach9

 Studies Using Paper-Pencil-Tests to Measure GPK Extensively15

 Studies Using Paper-Pencil-Tests to Measure a Small Segment of GPK31

 Studies Using a Videobased Stimulus for Testing Knowledge and/or Skills33

 Studies Rating the Demonstration of Teaching Skills39

 Phenomenological Studies Describing Teacher Knowledge42

 Studies describing differences between experts and novices.....44

PART 3: DISCUSSION OF CONCEPTUAL ISSUES46

 General Pedagogical Knowledge vs. Pedagogical Content Knowledge vs. Content Knowledge46

 Definition of Content47

 Cultural Sensivity51

 Assessment Formats.....53

PART 4. CONCLUSION AND RECOMMENDATIONS55

 Content Areas to be Covered and Types of Knowledge to be Assessed55

 Methodological Approach56

REFERENCES59

ANNEX A ABBREVIATIONS69

ANNEX B: TERM SEARCH RESULTS71

Tables

Table 1. Empirical Studies on GPK – Articles selected for Review – Results from Term Search and Chosen by the Author12

Figures

Figure 1: Dimensions and topics covered in the TEDS-M test of GPK (König et al., 2011, p. 191)17

Figure 2: Test design matrix (König et al., 2011, p. 191)18

Figure 3: Item example for GPK about.....19

Figure 4: Item example for GPK about.....19

Figure 5: Item example for GPK about.....	19
Figure 6: US future teacher’s response to item example 2 (König et al., 2011, p. 192)	20
Figure 7: Topics, Cognitive Processes, and Summaries of Items Examples (<i>italics</i>) covered in the test additionally used in the LEK study to measure GPK related to nurturing and school improvement (Seifert & Schaper, 2012, p. 185)	23
Figure 8: Scales assessing teacher quality used in a student questionnaire (Hopf, 2011, p. 90)	25
Figure 9: Topics and areas covered in the ETS assessment (ETS, 2013, p. 13) as well as number of items in the Grades 5-9 Test (ETS, 2011, p. 3)	30
Figure 10: Item example for the pedagogical skill to interpret a classroom situation (König et al., 2014, p. 82)	35
Figure 11: Topics and areas covered in paper-and-pencil-tests measuring GPK extensively	49
Figure 12: Topics and areas covered in other studies reviewed in Part 2	50
Figure 13: Pedagogy as ideas: theories, values, beliefs and justifications (Alexander, 2009, p. 17).....	51
Figure 14: Competence model proposed by König (2009, p. 248) using the TEDS-M instrument and integrating GPK content, cognitive demands, and explicit item difficulties (terminology of test items, complexity of cognitive processes)	58

SUMMARY

This expert paper aims at informing and making recommendations for designing an international assessment instrument related to teachers' general pedagogical knowledge (GPK). After providing some information on background issues and the policy context related to the state of research and the justification to assess teacher knowledge, a selection of some previously conducted studies will be thoroughly reviewed. Building on such research studies, conceptual issues such as the distinction of general pedagogical knowledge and pedagogical content knowledge, content of GPK, cultural sensitivity, and assessment formats with regards to teachers' GPK will be discussed. The paper will conclude with recommendations resulting from these considerations. Several issues relevant for a large-scale GPK test design will be taken into account and discussed throughout the report.

INTRODUCTION

What this Report is Aiming at

1. This report aims at informing what kind of suggestions and recommendations one could make for an international assessment instrument related to teacher general pedagogical knowledge (GPK), i.e., teacher generic knowledge and skills mainly related to teaching, but irrespective of the school subject and without regard to subject-specific nuances (for further details on how GPK can be defined, see Part 2). First, a background review is given and the policy context is outlined with regards to the state of research and justification for an assessment of teacher knowledge. Second, a selection of some previously conducted studies is thoroughly reviewed. Thirdly, building on such research studies, conceptual issues such as the distinction of general pedagogical knowledge and pedagogical content knowledge, the definition of content, cultural sensitivity, and assessment formats with regards to teachers' GPK are considered. Finally, recommendations are given resulting from these considerations. Also, open research questions, research desiderata, and the needs for upcoming research will be named.

2. Several issues relevant for a large-scale GPK test design will be taken into account and discussed throughout the report:

- What could be learned from previous studies measuring teachers GPK?
 - What do these studies have in common?
 - How do they vary in study design and in choices of research methods?
 - Which research questions have they raised and answered, and what are key findings?
 - Is there evidence in teachers' GPK supporting student learning?
- What are central conceptual issues coming to the front when teachers' GPK is to be measured?
 - How can GPK be separated from adjacent knowledge domains, especially Pedagogical Content Knowledge (PCK)?
 - Is consensus building about what is content of GPK when regarding how the studies conceptualised and operationalised GPK?
 - What are key aspects of defining and measuring GPK, including cultural sensitivity of pedagogy in an international context?
 - Which assessment formats have been used in order to assess teachers' GPK directly?
- What could be recommended on developing an international assessment of GPK and factors/variables to consider for assessment?
 - How can content areas within GPK be defined?
 - Which types of knowledge have to be accounted for, including specific and general aspects of methodological approaches?

3. As far as the author knows no review exists that deals with the field of the topic under study. There is a review of the state of research on the assessment of mathematics teacher knowledge across countries (Blömeke & Delaney, 2012), but this is subject-specific and does not deal with general pedagogical knowledge of teachers.

PART 1: BACKGROUND REVIEW AND POLICY CONTEXT

On the State of Research and the Justification for an Assessment of Teacher Knowledge

Introduction

4. Since a few years, a body of research on the knowledge of teachers and future teachers has been growing. This recent development can be interpreted as an answer to the research findings provided by international comparative studies like TIMSS and PISA that assess the achievement of students. Contrary to expectations, a number of countries have repeatedly not shown satisfying results. And even if some countries have had better results than others, the promotion of young people's learning is a common aim of all OECD countries.

5. To change this and assure a higher quality in education, various issues have been discussed. A significant discourse looks at the teacher quality and the quality of teacher education (OECD, 2005; Hattie, 2009; 2012; Schleicher, 2011). Then, the underlying assumption is that, due to a connection between the competences of teachers and their performance and teaching practice in the classroom, with an increase of teacher quality and quality of teacher education, teachers can provide high quality opportunities to learn at school and thus indirectly contribute to an increase in student achievement.

6. In this context, the measuring of the knowledge of teachers and future teachers is clearly essential. Research on teacher expertise conducted as early as in the 1980ies and 1990ies let us assume professional knowledge is a significant factor for effective teaching and thus promoting student attainment. But still there is the need to evaluate teacher knowledge as an outcome of teacher education systems and as a predictor for effective teaching and student attainment (e.g., Baumert et al., 2010). This is due to a lack of assessment tools that would allow countries to compare, analyse, and improve teacher quality and teacher education programmes and curricula on the basis of scientific criteria related to teacher knowledge.

State of Research

7. Most important aspects related to the state of research seem to be:

- there is a lack of empirical studies on (future) teachers' GPK (e.g., Wilson & Berne, 1999; König et al., 2011; Voss et al., 2011);
- some studies use distal indicators such as certification of teachers (e.g., Darling-Hammond, 2000), but actually do not directly test teacher knowledge as a predictor for teaching quality or student achievement (cf., e.g., Baumert et al., 2010);
- the few studies that actually test teacher knowledge focus on content knowledge and/or pedagogical content knowledge, mostly exemplified by mathematics teachers (e.g., Baumert et al., 2010; Ball, Thames & Phelps, 2008; Hill, Rowan, & Ball, 2005; Schmidt et al., 2007; 2011; Tatto et al., 2012); (general) pedagogical knowledge can be regarded as a neglected category (cf. Atjonen et al., 2011; König et al., 2011);
- the few studies that have been carried out to test teacher knowledge or to investigate teacher knowledge phenomenologically usually work with small sample size, i.e., clearly $n < 100$ (e.g.,

Bouas, 1996; Baer et al., 2011). Large-scale assessment of teacher knowledge has been rarely carried out and can be regarded as a new research type, especially when these studies compare knowledge internationally (cf., e.g., Klieme, 2012).

- as can be shown in this review, there is a tremendous research deficit regarding teaching quality or student achievement as an outcome of teachers' GPK. In fact, only two studies included in this report provide first findings on the possible influence of the pedagogical knowledge of pre-service teachers (Voss et al., 2011) and in-service teachers (Pflanzl et al., 2013; König & Pflanzl, under review) on facets of in-class teaching performance as rated by students taught by these teachers. Since even these innovative studies do not tell us about the effect of teachers' GPK on student achievement, the need for conducting such studies in the near future becomes obvious.

Justification for an Assessment of Teachers' General Pedagogical Knowledge¹

8. Important arguments to justify an assessment of pre-service and in-service teachers' (general) pedagogical knowledge seem to be at least twofold. It can be justified from the perspective of *in-service teachers* and the need for their ongoing *professional development*: What learning needs do teachers have? How can they precisely identify their learning needs? What criteria can be used to determine which knowledge they have to acquire during professional development? What scientific criteria are used to evaluate the effectiveness of professional development courses or programmes? The other perspective is related to *initial teacher education* and *pre-service teachers*: How should teacher education programmes and curricula be designed? What knowledge should teacher educators consider when teaching student teachers? What scientific criteria are used to evaluate the effectiveness of teacher education programmes and systems? What knowledge should teachers have when entering the teaching profession?

9. Teachers need generic knowledge for successful teaching, e.g. an "intellectual framework" for classroom management (Doyle, 1985; 2006; Shulman, 1987; Borko & Putnum, 1996) or, more generally, knowledge of pedagogical concepts, principles and techniques that is not bound by topic or subject matter (Grossman, 1990; Wilson, Shulman & Richert, 1987). Teachers draw on this knowledge and weave it into coherent understandings and skills when they deal with the learner and subject matter in the classroom (Shulman, 1986; 1987). One might argue that teaching is a profession in which professional knowledge should be defined more clearly in order to support systematic planning of teacher training curricula and programmes and to help teachers to identify their learning needs during professional development. This is even more important with regards to ill-structured knowledge domains such as pedagogy than, for example, with regards to structured knowledge domains such as mathematics. But instead, research shows teachers often report they think they have acquired their pedagogical knowledge primarily through 'trial and error' as well as through individual reflecting on their teaching practice (Hativa, 2000; Garrahy et al., 2005). The dominance of professional practice and pedagogical routines and the absence of the theoretical, academic input might result in conserving traditional teaching styles and habits at school that will not be enriched by new knowledge relevant for teaching such as new research findings from the learning sciences.

10. General pedagogical opportunities to learn form an essential part of initial teacher education systems in many countries worldwide (Schmidt et al., 2007; König, Tachtsoglou, Darge & Lünemann, 2014). Decisions that are made in the context of reforms of teacher education should be built on evidence-based ground (e.g., König et al., 2011, p. 188), especially with regard to general pedagogy as a component of teacher education programmes, broad claims about its uselessness, as well as about what future teachers

¹ This paper does not aim at coming to a consensus on the definition of 'teacher knowledge'. Building on widely used conceptualisations of teacher knowledge (such as Shulman, 1987), its scope is to review other research on this topic and how those researchers have defined teacher knowledge.

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 provide promising hypotheses, without empirical testing they have their limits in the process of improving teacher education (Larcher & Oelkers, 2004). “Looking at initial teacher education programs and their variation across institutions, regions, and countries presumably policy makers, institutions, curriculum designers, teacher educators, and pre-service teachers are in need of comparative data and of specific evaluative feedback” (cf., e.g., Pecheone & Chung, 2006). But there has been little research into initial teacher education in general pedagogy. For example in Europe, despite the Bologna reforms, heterogeneous policy decisions – unsupported by empirical data – have produced a situation in which the content and the characteristics of initial teacher education programmes are hardly comparable and lack evaluative feedback.

11. In this context, clearly the Teaching and Learning International Survey (TALIS) is a very important initiative, circularly collecting comparative data of several ten thousands in-service teachers in more than 20 countries worldwide. The survey focuses on issues such as vocational status, working conditions, and beliefs of teachers, and a central aim of TALIS is to identify correlations between background variables of teachers, their working conditions and beliefs, and self-reported teaching practices. Since TALIS produces high quality data on teachers, teaching, and teacher education on an international comparative level, its potential is obvious to integrate an assessment of teacher knowledge in a future cycle in order to meet core aspects of current research desiderata and to provide data highly relevant for educational policy.

PART 2: A REVIEW OF SOME PREVIOUSLY CONDUCTED STUDIES

Methodological Approach

Introduction to the Section

12. In the following we will outline how we have carried out the review of the studies, how they were selected and which criteria we applied when describing them. Part 2 primarily aims at a *description* of these studies, while part 3 has more an *analytical* point of view. In Part 4, we draw conclusions and give recommendations, so Part 4 is much more *interpretative*.

How Studies were Selected

13. In May 2013, a term search using the following three databases was conducted: *ERIC*, *PsycInfo*, and *Web of Science*. Search term was “*pedagogical knowledge*” (between quotation marks) that needed to be mentioned in the abstract, keywords, and/or title.²

Using the following five criteria, the number of results was reduced:

- (1) Published in peer-reviewed journal (i.e., articles)
- (2) From 1998 to 2013
- (3) Written in English
- (4) Explicit focus on pedagogical knowledge and not on:
 - *pedagogical content knowledge* (PCK). According to the frequently cited classification of teacher knowledge provided by Shulman (1986; 1987), PCK is “that special amalgam of content and pedagogy that is uniquely the province of teachers, their own special form of professional understanding“ (Shulman, 1987, p. 8). There is broad consensus today PCK is bound to the specific domain and not irrespective of the subject;
 - *curriculum knowledge; content knowledge*. Also these teacher knowledge categories were introduced by Shulman (1987, p. 8). They are related to a specific domain. Content knowledge is related to the particular discipline and independent from its pedagogical application in teaching. Curriculum knowledge is related to the materials that structure a school subject;
 - *technological pedagogical content knowledge* (TPCK); *technological pedagogical and content knowledge* (TPACK); *pedagogical technology integration content knowledge*

² “General pedagogy” or “pedagogy” sometimes denotes the area in a broader sense, but as far as the review of literature conducted here shows, then authors do not at all (or do not necessarily) think of teacher knowledge as an individual disposition. Instead, other issues are considered, e.g., opportunities to learn in teacher education, pedagogy as an academic discipline, or pedagogy in general but not specifically related to the teaching profession etc. Since this paper’s objective is to review studies that specifically deal with the topic of assessing teachers’ GPK, that kind of literature was excluded.

(PTICK). Mishra and Koehler (2006; 2009) have expanded teacher knowledge classifications with regards to application of new media and other technology use in the classroom. Also, there is a clear relation to teaching subjects;

- *context knowledge*; again, Shulman (1987, p. 8) defines this as knowledge “ranging from the workings of the group or classroom, the governance and financing of school districts, to the character of communities and cultures“. Since this paper focuses on tasks closely related to teaching in particular, we do not account for this knowledge here;
- *biographical experiences*; studies focusing on what is called biographical experiences may appear as reflective papers (e.g., on a single teacher’s expertise), but they do not bring substantial evidence from empirical data. Studies reporting about biographical experiences use a methodology that is difficult to integrate in this paper.

(5) only articles that report on empirical (quantitative and/or qualitative) research, thus excluding purely conceptual papers and discussion papers.

14. This term search produced a total number of 69 results (39 in *ERIC*, 18 additional in *Psychinfo*, and 12 additional in *Web of Science*). All 69 results are listed in Annex B at the end of this report. The five criteria were applied again while reading the abstracts. The author of this paper commented on each of the articles and decided to select or deselect (see the two right columns in the table in Annex B).

15. After applying these criteria it turned out only 11 research articles have a clear focus on “pedagogical knowledge”, are empirical studies (mostly qualitative, some quantitative), and are related to school teachers (either pre-service or in-service). Three articles (Choy et al., 2011; 2012; O’Hara et al., 2008) reported about research conducted with *self-report scales* and thus were excluded (see next section for further details). Another one clearly had a different target group than school teachers (Hativa, 2000), and another one was related to rating teacher performance (Brown et al., 2006). Although the remaining six articles (Atjonen et al., 2011; Garrahy et al., 2005; König et al., 2011; König & Rothland, 2012; Zohar, 2002; 2006) provide a variety of research in terms of methods, authors etc., four articles come from two research teams or approaches (on the one hand: König et al., 2011; König & Rothland, 2012; on the other hand: Zohar, 2002; Zohar, 2006), so in the end the six articles only stand for four different approaches to do research on teachers’ GPK.

16. From the results of this search one can conclude that (general) pedagogical knowledge of teachers has been rarely investigated in comparison to content knowledge and pedagogical content knowledge. If the term “pedagogical content knowledge”, for example, is included in the search of *ERIC* database, 276 results are provided (instead of 39 results we received here). Many articles that were included in search results are in fact related to pedagogical *content* knowledge, so these were excluded when the criteria defined above were applied. Another way to find a rather small selection of articles is to search for “general pedagogical knowledge” instead of “pedagogical knowledge”. But this search was not evaluated for this review.

17. Because term search using well-known databases resulted in a very small number of research articles only, in addition to this and to enlarge the review, the author of this paper decided to include some more studies that he considered to be relevant for this review although they were not indicated in the term search. The following aspects were guiding his decisions:

- empirical studies measuring teachers’ GPK using a standardised and elaborated test instrument that were not listed in the term search (e.g., because they were not published in English or because they

were in the status of being submitted/under review/in press or published before 1998) but well-known to the author (Bouas, 1996; König, Blömeke, Klein, Suhl, Busse & Kaiser, 2014; König, Buchholtz & Dohmen, under review; König & Lebens, 2012a; Kunina-Habenich et al., 2013; Seidel, Blomberg & Stürmer, 2010);

- studies in which the demonstration of teaching skills are focused on (Pecheone & Chung, 2006; 2007; König, Buchholtz & Dohmen, under review), but which are not (only) related to rating teacher performance in class (see the next section for further details on this);
- phenomenological (qualitative) studies that do not actually test teacher knowledge, but that aim at exploring and describing knowledge of teachers, especially those working with expert and novice teacher comparisons. Most of them were published before 1998 (see, e.g., the overview given by Hogan et al., 2003), but it seems to be important to provide at least summarised information in this paper regarding the question what they can contribute to the issue of assessing teachers' GPK.

18. Table 1 provides a summary of the key publications that will be referred to in the following. Quite often, one publication is related to a larger project, therefore representing this project or even follow-up projects that work with the same research instrument. If this is the case, some more publications will be named when the study is described.³

³ When the term-search was conducted as outlined here, no specific study from the Asian context could be found. However, at least with the comparative study "TEDS-M" (a study included and described in the following section "Studies Using Paper-Pencil-Tests to Measure GPK Extensively") there is Asian research insofar as GPK of future Taiwan teachers is examined.

Table 1. Empirical Studies on GPK – Articles selected for Review – Results from Term Search and Chosen by the Author

(Number in the left column refers to results from term search (see Annex B), a star marks that this publication was added by the author)

No	Reference	Study name	Sample size	Target group	Country /countries involved	Cross-national analysis	National regional	Name or short description of test instrument or construct	Research method
Studies Using Paper-Pencil-Tests to Measure GPK Extensively									
64	König et al., 2011	TEDS-M	1,743	Future teachers at the end of training	U.S., Germany, Taiwan	x		General pedagogical knowledge for teaching (TEDS-M instrument)	Paper-pencil-test, quantitative, IRT-scaling
46	König & Rothland, 2012	DIDAKTUM	130	Student teachers from one university	Germany		x	General pedagogical knowledge for teaching (TEDS-M instrument)	Paper-pencil-test, quantitative, IRT-scaling, longitudinal data analysis
*	König & Seifert, 2012; König, 2013	LEK	261 Panel sample, two occasions of measurement	Future teachers during their first two years of training	Germany		x	General pedagogical knowledge for teaching (TEDS-M instrument) + educational knowledge test related to topics of school development and nurturing students	Paper-pencil-test, quantitative, IRT-scaling
*	König et al., 2013	EMW	6,601	First year student teacher cohort	Germany, Austria, Switzerland	x		General pedagogical knowledge for teaching (TEDS-M instrument)	Paper-pencil-test, quantitative, IRT-scaling
*	Pflanzl et al., 2013	SKILL	55 teachers 877 students	Beginning in-service teachers and their students	Austria		x	General pedagogical knowledge for teaching (TEDS-M instrument) + student ratings on teacher quality	Paper-pencil-test, quantitative, IRT-scaling; student survey
*	Kunina-Habenicht et al., 2013	BILWISS	3,273	Pre-service teachers	Germany		x	“Educational-Scientific Knowledge” (very broad conceptualisation of general pedagogical knowledge)	Paper-pencil-test, quantitative
*	Voss et al., 2011	COACTIV-R	746	Future teachers (second phase)	Germany		x	Pedagogical-psychological knowledge	Paper-pencil-test, video-vignettes, quantitative, SEM
*	ETS, 2013	PRAXIS II		Future teachers at the end of training	U.S.		x	general PPK (principles of learning and teaching)	Paper-pencil-test, text-vignette (case studies), quantitative
Studies Using Paper-Pencil-Tests to Measure a Small Segment of GPK									
*	Bouas, 1996		53	Pre-service teachers	A mid-western university in the US		x	teachers’ knowledge of cooperative learning	10 MCR-Items to test future teachers’ knowledge of cooperative learning as a teaching strategy; interviews; data were triangulated
*	Scituo, Terjesen, & Frank (2000)	KADDS	149	Elementary school teachers from six New York area public schools	New York area, USA		x	teachers’ knowledge of Attention Deficit Disorder Scale (KADDS)	36-item rating scale, quantitative

Studies Using Videobased Stimulus for Testing Knowledge and/or Skills									
*	König et al., 2014	TEDS-FU	171 panel sample, two occasions of measurement	middle school mathematics teachers	Germany		x	General pedagogical knowledge for teaching (TEDS-M instrument) + General pedagogical video-vignettes measuring noticing and interpreting of classroom events	Paper-pencil-test, video-vignette test, quantitative, IRT-scaling
*	König & Lebens, 2012a	CME	108	Pre-service teachers, in-service teachers	Germany		x	Classroom Management Expertise measured by video-vignette instrument	Video-vignette test, quantitative, IRT-scaling
*	Seidel, Blomberg & Stürmer, 2010	OBSERVE			Germany		x		Video-vignette test, quantitative
54, 55	Zohar, 2004; see also Zohar, 2002; 2006	Thinking in Science Classroom (TSC)	83	Israeli junior and high school science teachers	Israel		x	Mixture of pedagogical knowledge and beliefs, unclear construct, very explorative	Qualitative, video-based, explorative
Studies Rating the Demonstration of Teaching Skills (without rating the performance in class)									
*	Pecheone & Chung, 2006	Performance Assessment for California Teachers (PACT)	395 (during 2002-2003 pilot), 628 (during 2003-2004 pilot)	Future teachers to be certified (i.e., at the end of training)	California, USA		x	“Teaching Events” (TE) consisting of tasks: to plan and teach a learning segment; to videotape and analyse one’s instruction; to collect student work and analyse student learning; to reflect on their practice	Rating the Demonstration of Teaching Skills, no knowledge test
*	König, Buchholtz & Dohmen (under review)	PLANVOLL	106	future teachers in their last year of training, two occasions of measurement	Berlin, Germany		x	Lesson plan analysis, dealing with heterogeneous learning of students	11 dichotomous indicators to analyse lesson plans, IRT-scaling, longitudinal data analysis
Phenomenological Studies Describing Teacher Knowledge									
44	Atjonen et al., 2011		2,351	in-service teachers fourth grade, 7 th to 9 th grade	Finland		x	Pedagogical key principles, obstacles, and success	No testing; qualitative analysis of open-end items, frequency distribution and correlative analysis
8	Garrahy et al., 2005		20	in-service physical education teachers	USA		x	Classroom management; no such construct, but a description of „knowledge content“ evolved from qualitative interview data	Interview data to identify key aspects of teachers’ cognitions about classroom management
*	Sabers, Cushing & Berliner, 1991			Summarised information only following Hogan et al., 2003			x	Classroom management	Expert-novice comparison, qualitative
*	Clarridge & Berliner, 1991								
*	Swanson et al., 1990								
*	Sánchez, Rosales, & Canedo, 1999								
*	Allen & Casbergue, 1997								
*	Gonzales & Carter, 1996								
*	O’Connor & Fish, 1998								

Studies not included

19. Since this paper aims at informing and making recommendations for designing an international assessment instrument related to *teachers' general pedagogical knowledge* (GPK), we narrowed it down to a selection of certain studies. At the same time, certain studies or certain types of studies were deliberately not selected.

20. There is a substantial number of studies capturing aspects of GPK with self-reports of future teachers (e.g., Oser & Oelkers, 2001; Choy et al., 2011; 2012). Although some of these studies are somehow related to teachers' GPK, they do not test knowledge. Instead, they ask (future) teachers to respond to Likert-scales in order to measure self-perceptions of knowledge or self-perceptions of competences. These studies were not included into this review. One major reason is that this methodological approach has been critically scrutinised with regards to validity. Researchers argue that self-reports about one's own knowledge are influenced by personality features (such as, e.g., self-efficacy), but they do not actually measure knowledge (see König, Kaiser, & Felbrich, 2012).

21. Similarly to studies using self-reports, there are studies that use rating scales with which a rater is asked to assess (future) teachers' performance in class (e.g., Brown et al., 2006). These studies were not included into the report either, since they do not measure knowledge directly. Instead, they are much more related to the research on effective teaching or instructional quality which investigates quality aspects of teaching in class and the performance of teachers (e.g., how they behave in class). The PACT (Pecheone & Chung, 2006; 2007) is a borderline case, since it includes ratings of teachers' performance as well. But because it includes other components (such as analysis of lesson plans) it will be included into this review.

How Studies are Grouped for Describing them in this Part

To describe the studies in this part, they were assigned to five groups:

- studies Using Paper-Pencil-Tests to Measure GPK Extensively;
- studies Using Paper-Pencil-Tests to Measure a Small Segment of GPK;
- studies Using Video-based Stimulus for Testing Knowledge and/or Skills;
- studies Rating the Demonstration of Teaching Skills (without rating the performance in class);
- phenomenological Studies Describing Teacher Knowledge.

22. First, we will describe quantitative empirical studies that use paper-and-pencil-tests to measure teacher GPK. There are obvious differences between studies that attempt to measure GPK extensively (e.g., often combined with more complex scaling analysis) and studies that focus on a very specific aspect only (e.g., cooperative learning), so this will be mirrored by the first two groups.

23. Second, we will describe quantitative studies using video-based stimuli for testing knowledge. They usually make also use of the paper-and-pencil approach, but compared with purely paper-and-pencil studies, they intend to measure knowledge that is more of a situated nature (Putnam & Borko, 2000). So these studies are assigned to a third group. In this group, we will also report about qualitative studies that did not test knowledge, but that used video-based stimuli and open-response questions and analysed teachers' responses with qualitative methods.

24. When trying to go beyond paper-and-pencil testing or using video-based stimuli, we can identify studies that intend to rate demonstrations of teaching skills. As outlined before, this review does neither deal with ratings of teacher performance in class nor with video-taped performance, since this would be subject of video-studies such as reviewed by Klette and Bergem (2012) in the TALIS Video Study Report. Instead, we focus on studies that, for example, analyse lesson plans as an approach of competence measurement. This forms the fourth group of studies.

25. Finally, although not primarily focused here, we consider it important to summarise findings from phenomenological studies, especially a selection of studies from the research on teacher expertise, so this will be the last group of studies. This is important, since these studies provide a sort of background information for this review.

26. When we describe the main aspects of each study reviewed in this paper, we follow a certain structure: First, *background information and summary of study design* is provided. Second, the *test instrument and the choice of technical solutions* are described. *Major research questions* and *key findings* are outlined. Finally, conclusions from the particular study are drawn with regards to the *use for this review*. Information provided in this section will be the basis for the discussion and recommendations of the succeeding parts.

27. Since some of the studies reviewed are quite complex, only a selection of aspects, research findings etc. can be reported here. If possible, we will include in the sections on research questions and key findings information about the question whether the instrument used in the specific study was used to predict student outcomes or any other analyses conducted for relating teachers' results to student achievement. However, there are only very few studies that have conducted these analyses (see Pflanzl, Thomas & Matischek-Jauk, 2013; König & Pflanzl, under review; Voss, Kunter & Baumert, 2011).

Studies Using Paper-Pencil-Tests to Measure GPK Extensively

Introduction to the Section

28. In the following we will describe studies that fulfil the following criteria: They focus on General Pedagogical Knowledge (GPK) of pre-service or in-service teachers, they use a standardised measurement of GPK using paper-pencil-test instruments, and they use a test design to measure GPK extensively (i.e., they measure various facets or dimensions of GPK by making use of more complex scaling analysis).

TEDS-M (König, Blömeke, Paine, Schmidt & Hsie, 2011)

Background Information and Summary of Study Design

29. The *Teacher Education and Development Study – Learning to teach Mathematics* (TEDS-M) is a comparative study of teacher education with a focus on the preparation of teachers of mathematics at the primary and lower secondary levels. It is the first study of the IEA in tertiary education as well as the first international large-scale assessment of future teachers that works with representative samples (see for details, Tatto et al., 2008; 2012). 17 countries worldwide participated in TEDS-M. The TEDS-M target population are future teachers in their final year of training before they are fully certified to teach in schools as ordinary school teachers. The central component of the TEDS-M study is to measure the professional knowledge of future teachers. As the core of the International Questionnaire only measures future teachers' knowledge related to CK and PCK, some countries – Germany, Taiwan, and the USA – decided to participate in a national option measuring future teachers' GPK that was developed under the leadership of the German TEDS-M team (see for details, König, Blömeke, Paine, Schmidt & Hsieh, 2011).

30. TEDS-M was built on previous research, particularly conducted in the study *Mathematics Teaching in the 21st Century* (MT21; Schmidt et al., 2007). In MT21, research on measuring GPK was started (Blömeke et al., 2008) and then intensified in further research conducted in TEDS-M (see for details, e.g., König & Blömeke, 2009). Since principle ideas of the TEDS-M measurement go back to the pioneer work in MT21, a more detailed description of the TEDS-M test instrument will be focused here.

31. Besides TEDS-M, there are several studies that have applied the test instrument, mostly conducted in the German speaking countries. Four of them will be described in the next four sections:

- Following our term search results (see previous section) we will report on the so-called DIDAKTUM-Study (König & Rothland, 2012; see next section).
- Another study, the so-called LEK-study, carried out in Germany applied the test instrument to analyse future teachers' GPK when entering teacher training and after two years. Future teachers from four universities were included into this study and they were tested twice in order to analyse knowledge change during initial teacher education. Importantly, besides the TEDS-M measurement instrument, another test was used to measure GPK extensively in the area of educational knowledge. So this study seems to be relevant for this review, too.
- Then, in autumn 2011, an international study was launched, the so-called EMW-study that assesses the GPK of future teachers in the three German-speaking countries Germany, Austria, and Switzerland. It will also be reported on here, since it works with large sample size ($n > 6,000$) and since it is one of the very few international-comparative studies that measure GPK in a cross-national design.
- Finally, the SKILL-study is presented (Pflanzl et al., 2013; König & Pflanzl, under review) with regards to findings on the relationship between in-service teachers' GPK score ($n = 55$) and student ratings on teacher quality ($n = 877$). Mentioning this study seems to be important, since it provided further insight into the validity of the TEDS-M measurement instrument.

Test Instrument and Choices of Technical Solutions

32. The theoretical framework of GPK developed in the context of TEDS-M is structured in a task-based way and explicitly not according to the formal structure of general pedagogy as an academic discipline. Furthermore, instruction is conceptualised as the core activity of teachers and serves as a heuristic to select topics and cognitive demands of GPK. Findings from instructional research and didactics were combined to conceptualise GPK for teaching as is shown in Figure 1 (for details, see König et al., 2011): Four dimensions of GPK are considered highly relevant with respect to the target group of future teachers. Teacher education is regarded as effective if future teachers in the last year of their training have 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 assess students ("assessment").

Figure 1: Dimensions and topics covered in the TEDS-M test of GPK (König et al., 2011, p. 191)

Test 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
	- effective use of allocated time/ routines
Adaptivity	- strategies of differentiation
	- use of a wide range of teaching methods
Assessment	- assessment types and functions
	- evaluation criteria
	- teacher expectation effects

33. Apart from the task-based dimensions and topics of GPK shown in Figure 1, dimensions of cognitive processes were additionally defined to describe the cognitive demands on future teachers when they respond to test items. This was done, because it was assumed that teachers' GPK is multidimensional and thus of different quality. Following Anderson and Krathwohl's elaborate and well-known model (2001), with the TEDS-M test instrument measuring GPK, three cognitive processes were distinguished which summarise the original six processes: recalling, understanding/analysing, and generating.

- *Recalling.* Future teachers have to retrieve information from long-term memory in order to respond to a test item. Test items of that type challenge future teachers to give an example for a definition, to recite elements of a phenomenon, a term, a concept, or to identify a term or a concept.
- *Understanding/Analysing.* In order to respond to a test item of this type future teachers also have to retrieve information from long-term memory, but, moreover, they have to link that information with a problem outlined by the test item. So they have to describe or explain a phenomenon or a concept; or they are asked to compare, categorise, assign or interpret a phenomenon, a situation or one or several general terms.
- *Generating.* To respond to items of the third dimension of cognitive processes future teachers have to generate concrete strategies concerning how they would solve a typical classroom situation problem which includes evaluating this situation. Again, retrieving information from long-term memory might be helpful, but, moreover, that knowledge has to be linked with classroom situation experience. Future teachers are asked to explicate practical knowledge that can be described as propositional mental representation. This item type reflects the need to measure GPK that is of a situated nature (Putnam & Borko, 2000).

34. The definitions of cognitive processes suggest GPK is of different quality. Furthermore, they can be related to the well-known differentiation between declarative knowledge ("knowing that...") and procedural knowledge ("knowing how..."), a distinction that is typical for the research on teacher knowledge (Anderson & Krathwohl, 2001). While the first two dimensions of cognitive processes

(recalling, understanding/analysing) predominantly test declarative or conceptual knowledge⁴, items of the third dimension (generating) also intend to measure procedural knowledge.

35. The test was designed to include variation in cognitive processes, because it was assumed that for a fully understanding, GPK has to be differentiated according to such processes. Major reason for this is the research on teacher expertise, according to which a novice teacher starts with the acquisition of declarative knowledge (e.g., about principles of instruction and classroom management), and then he or she starts to apply that knowledge when teaching students. Linking teaching experience to pedagogical knowledge (e.g., through reflection), a novice teacher progresses to the stage of the “advanced beginner” (see for details, Berliner, 2001; 2004). In fact, further research findings (König, 2013) showed (see for more details, the description of the LEK-study below) that future teachers at the end of training outperformed future teachers who had studied for two years, whereas they in turn outperformed future teachers who had just entered initial teacher education. Moreover, when analysing subscales of the test measuring cognitive dimensions of GPK, as would be expected declarative-conceptual knowledge (measured by cognitive dimensions “recall” and “understand/analyse”) are gained predominantly during theoretical opportunities to learn, whereas future teachers who had additionally passed through the practical in-school opportunities to learn perform much better on the test subscale “generate”. Therefore, the differentiation into cognitive processes contributes both to the research on teacher expertise and to the research on the differential influence of opportunities to learn during initial teacher education.

Figure 2: Test design matrix (König et al., 2011, p. 191)

	Recall	Understand/ analyse	Generate
Structure			
Motivation/classroom management			
Adaptivity			
Assessment			

36. In TEDS-M, topics of GPK and cognitive demands made up a matrix, which served as a heuristic for item development (see Figure 2). For each cell, a subset of items was developed. Several expert reviews 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. Moreover, their research had to be related to the topic of teacher knowledge and they had to be at least PhD candidates. In contrast to this first expert review, experts that participated in the second and following reviews, which aimed at selecting items for the final test instrument according to specific criteria or at validating the test instrument respectively, had to endow a university chair with a specialisation on research about teacher knowledge. Based on these review 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 (see, for details, König et al., 2011).

⁴ De Jong and Ferguson-Hessler (1996, p. 107) state: “*Conceptual knowledge* is static knowledge about facts, concepts, and principles that apply within a certain domain. Conceptual knowledge functions as additional information that problem solvers add to the problem and that they use to perform the solution. In our earlier work we used the term *declarative knowledge*.”

37. Three item examples (see Figures 3, 4, 5) illustrate the GPK test and the heuristic used to conceptualise GPK (see Figure 1 and 2). 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 analyse 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”.

Figure 3: Item example for GPK about “motivation” and “analyse” (König et al., 2011, p. 192)

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 <u>row</u>.</i>		
	intrinsic motivation	extrinsic motivation
A student learns before a test in mathematics, because he/she...		
A.	<input type="checkbox"/>	<input type="checkbox"/>
B.	<input type="checkbox"/>	<input type="checkbox"/>
C.	<input type="checkbox"/>	<input type="checkbox"/>
D.	<input type="checkbox"/>	<input type="checkbox"/>
E.	<input type="checkbox"/>	<input type="checkbox"/>

Figure 4: Item example for GPK about “structure” and “generate” (König et al., 2011, p. 192)

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

Figure 5: Item example for GPK about “assessment” and “generate” (König et al., 2013, p. 1006)

<p>At the end of a lesson, how can students receive useful feedback about their learning during the lesson? Give three different methods you find exceptionally useful.</p>

38. The second and third item examples are open-response items (see Figure 4 and 5). Regarding the item shown in figure 4, 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 analyse 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. Similarly, when responding to the test item shown in figure 5, future teachers had to “generate” different methods that would be useful when providing feedback to their students at the end of a lesson.

Figure 6: US future teacher's response to item example 2 (König et al., 2011, p. 192)

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

39. For the open-response items, coding rubrics were developed and reviewed by experts on teacher education. First, coding instructions were developed in a complex and extensive interplay of deductive (from the theoretical framework) and inductive approaches (from empirical teacher responses). In a pilot phase, codes from several independent raters were discussed in detail and coding instructions were carefully 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.

40. In TEDS-M, all questionnaires were coded on the basis of the coding manual. Two raters coded the answers independently of one another. As a measure of consensus and internal consistency Cohen's Kappa was estimated (Jonsson & Svingby, 2007). It ranges from .80 to .99 with an average of $M = .91$ ($SD = .07$). This can be regarded as a good result. If conformity of raters was lacking, an agreement between the two raters was obtained in collective discussion, calling on a third rater if necessary.

41. After having established reliable *coding* schemes, *scoring* strategies for complex open-response items such as the two shown in Figure 4 and 5 were developed in order to decide which codes could be rewarded and which could not because they did not seem to be appropriate. Again, experts had to agree on appropriate answers, which would sufficiently reflect expected outcomes of teacher education. Illustrating this strategy with the test item shown in Figure 4, codes were scored as appropriate if they addressed four criteria: "context" of the lesson (e.g., prior knowledge of students), "input" (e.g., objectives of the lesson), "process" (e.g., teaching methods used), and "output" of the lesson (e.g. student achievement). The excerpt of an original answer given by a US future teacher in the TEDS-M survey (see Figure 6) is a good example for these four criteria. For the test item in Figure 5, codes were scored as appropriate if they addressed three criteria: "teacher-centered" (e.g., oral inquiry by the teacher), "interactive" (e.g., discussion of work results), and "student-centered" (e.g., self-evaluation by students).

Major Research Questions

42. Future teachers' TEDS-M data from the GPK test have been analysed with respect to several research questions. One asked for the mean level of achievement of future teachers from the US, Taiwan, and Germany in order to compare the outcomes of teacher education in these three countries. Moreover, hypotheses related to the multidimensional structure of GPK as proposed by the theoretical framework were tested. And finally, effects of teacher education programmes' opportunities to learn on future teachers' GPK were analysed.

Key Findings

43. The findings revealed that future teachers from the US were significantly outperformed by future teachers in Germany with regard to the overall GPK test score (see for details König et al., 2011; König, Blömeke, Paine, Schmidt, Hsieh, 2014). The difference of nearly 1.5 standard deviations was very large, meaning that there was almost no overlap between US teachers and German teachers. Most of the lowest

scoring teachers from Germany did better than most of the highest scoring teachers from the US. By contrast, no differences could be found between Germany and Taiwan.

44. TEDS-M data analysis showed GPK is a multidimensional construct. Findings from *Item-Response-Theory* (IRT) scaling analysis show that it is legitimate to assess GPK as a one-dimensional construct, which is important, for example, when a summary of teacher performance across countries is needed. However, several indicators revealed that it is reasonable to distinguish between various topics of GPK and cognitive processes, as were developed in the theoretical framework outlined above. So future teachers' performance can accurately be described, and strengths and weaknesses of a country's teacher education programme outcomes can be evaluated in more detail. Technically speaking, multidimensional IRT models, specifying the sub dimensions of the theoretical framework, turned out to fit the data significantly better than a one-dimensional model did, suggesting only one general factor underlying the future teachers' responses (see for details, König et al., 2011; König, Blömeke, Paine, Schmidt, Hsieh, 2014). Only because of this differentiation, it was possible to identify the specific strength of future teachers of the US: Although they were outperformed in the international comparison by the other two countries, when looking at ipsative measures, their relatively high scores in test items measuring the cognitive process of generating fruitful pedagogical strategies came to the front and thus shed a positive light on the US initial teacher education (see for details, König et al., 2011).

45. When asking for possible effects teacher education programmes have on the acquisition of GPK during initial teacher education, multilevel modeling revealed significant positive effects of the number of topics studied in general pedagogy with regard to the GPK of future primary teachers in Germany – that is, the more opportunities to learn a future teacher had during her teacher education, the better she did on the TEDS-M test (Blömeke & König, 2011). Continuing the examination in this field, additional analysis showed that practical in-school experience served as a relevant opportunity to learn, too, and this with respect to future primary teachers in Germany and the US (König & Blömeke, 2012): Future teachers who had an appropriate balance between teaching and being supported by mentors and who had opportunity to reflect on and improve their teaching to a relatively large extent outperformed other future teachers who either had less support from mentors or who had gained only little teaching experience.

Use for this review

46. Because TEDS-M was led by the IEA, the testing of GPK across countries in TEDS-M, although carried out as a national option, fulfills high standards of quality. TEDS-M is the only large-scale assessment with representative samples from three culturally very different countries giving insight into the GPK of future teachers at the end of their training. Since no other comparative study of this kind has been carried out yet, TEDS-M is an important milestone for the purpose of this paper, e.g., regarding the theoretical conceptualisation of GPK for comparative analysis or technical solutions applied to solve challenges of cross-national surveys.

47. The following four studies to be presented (DIDAKTUM, LEK, EMW, and SKILL) have in common the application of the TEDS-M instrument measuring GPK. So they do not provide new technical solutions in terms of measuring GPK, but instead they give in-depth insight into the quality of the TEDS-M instrument. Since this is relevant for a better understanding of the TEDS-M instrument, these studies are discussed below.

DIDAKTUM (König & Rothland, 2012)

Background Information and Summary of Study Design

48. The DIDAKTUM study (*Didaktisches Wissen und berufliche Motivation von Lehramtsstudierenden / Didactic Knowledge and Vocational Motivations of Student Teachers*) was carried out from 2009 to 2010. It applied the TEDS-M instrument measuring future teachers' GPK and specifically focused on the relationship between teaching motivations and GPK as central constituents of teacher professional competence (Baumert & Kunter, 2006). Using longitudinal data of 130 pre-service teachers from one university in Germany, effects of pre-service teachers' motivations for choosing the teaching professions on their GPK are examined.

Test Instrument and Choices of Technical Solutions

49. See the previous section on TEDS-M for a description of the test instrument used in this study. Teaching motivations were measured via the German version of the FIT-Choice scale (*factors influencing teaching as a career choice*; Watt et al., 2012). It is fully documented in König and Rothland (2012).

Major Research Questions

50. The relationship between future teachers' teaching motivations and their GPK is investigated. The authors examine the effects of student teachers' motivations for choosing teaching on their general pedagogical knowledge; findings on this give further insight into GPK as a motivational outcome.

Key Findings

51. Evidence is provided that intrinsic motivation is positively correlated, whereas extrinsic motivation is negatively correlated with GPK at the first occasion of measurement. However, unexpectedly, extrinsic motivation has a positive effect on learning gain, whereas intrinsic motivation has not. Besides, findings show that there is a large learning gain over the course of only one year (König & Rothland, 2012, p. 303), indicating the curricular validity of the measurement instrument with regards to initial teacher education in Germany.

Use for this review

52. The study provides further insight into the TEDS-M measurement instrument testing GPK, since it provides empirical findings from correlational and longitudinal data analysis. However, the study is limited to initial teacher education at one university of Germany.

LEK (König & Seifert, 2012; König, 2013)

Background Information and Summary of Study Design

53. Since TEDS-M proliferates an understanding on what future teachers know in the final year of their training, the LEK study (*Längsschnittliche Erhebung pädagogischer Kompetenzen von Lehramtsstudierenden / Longitudinal Survey of Student Teachers' Pedagogical Competencies*) was carried out from 2008 to 2010 (see for details, König & Seifert, 2012). The TEDS-M test measuring GPK was applied to future teachers at different teacher education stages (the beginning, after 2 years) and thus to compare their performance with TEDS-M data (see for details, König, 2013).

54. Moreover, the LEK study made use of another test measuring GPK (see for details, Seifert & Schaper, 2012; Seifert & König, 2012). While the TEDS-M test measuring GPK is related to the teachers'

central task of teaching, the other test is related to two more tasks, namely nurturing students and a teacher's contribution in school organisation and improvement. However, while the TEDS-M test was conceptualised for international comparisons, the other test's conceptual focus is primarily to the situation in Germany where teachers' tasks are classified into teaching (including assessment), nurturing, and school improvement (KMK, 2004).

Test Instrument and Choices of Technical Solutions

55. See the previous section on TEDS-M for a description of the test instrument used in this study. The other test instrument related to teachers' tasks of nurturing students and school improvement is to be described here in more detail. Similar to the TEDS-M instrument, it is a paper-pencil test with about 30 test items allowing standardised measurement of teacher knowledge. Its theoretical framework works out the relevance of these two teachers' task in addition to the core task of teaching (see for details, Seifert & Schaper, 2012; Seifert & König, 2012). The rationale of the test instrument is summarised in figure 7 (translated from the figure in Seifert & Schaper, 2012, p. 185).

Figure 7: Topics, Cognitive Processes, and Summaries of Items Examples (italics) covered in the test additionally used in the LEK study to measure GPK related to nurturing and school improvement (Seifert & Schaper, 2012, p. 185)

	cognitive processes	to reproduce knowledge	to apply knowledge	to judge, to decide
topics				
nurturing and educating students		<i>to assign statements into a model of communication</i>	<i>to link examples with mechanisms of learning</i>	<i>to judge teaching situations with psychological learning theories</i>
school improvement		<i>to know central tasks of the teaching profession</i>	<i>to assign assessment criteria to educational goals</i>	<i>to judge on evaluation objectivity</i>

56. Similar to the TEDS-M test instrument, dimensions of cognitive processes were additionally defined to describe the cognitive demands on future teachers when they respond to test items. In general, two broader categories can be used here (see for details, Seifert & König, 2012): The test measures declarative-conceptual knowledge primarily with the test items under "to reproduce knowledge", while procedural knowledge is tested using items of the other two categories ("to apply knowledge", "to judge, to decide"). So the test instrument makes use of the well-known differentiation between declarative knowledge ("knowing that...") and procedural knowledge ("knowing how..."), a distinction that is typical for the research on teacher knowledge (Anderson & Karthwohl, 2001). IRT-scaling analysis was applied to examine the quality of the test (see for details, Seifert & Schaper, 2012).

Major Research Questions

57. The authors of the LEK study asked whether future teachers gain GPK during the course of their training and whether theoretical and practical opportunities to learn have an effect on different facets of future teachers' GPK. With regards to the measurement of GPK related to teaching (using the TEDS-M instrument), data from TEDS-M and LEK were linked to analyse differences in GPK between future teachers at different stages (beginning, two years after, end of training). With regards to GPK related to the teachers' tasks of nurturing students and school improvement using the other test instrument) data analysis was limited to two stages only (beginning, two years after).

Key Findings

58. Findings show that the more advanced future teachers are in the course of their initial teacher education, the better they perform in the TEDS-M test measuring GPK. When analysing subscales of the test measuring cognitive dimensions of GPK, declarative-conceptual knowledge (measured by cognitive dimensions “recall” and “understand/analyse”) was gained predominantly during the theoretical study of the German initial teacher education (first phase), whereas future teachers who had additionally passed through the practical second phase performed much better on the practical knowledge test subscale (measured by the cognitive dimension “generate”).

59. Regarding GPK related to nurturing of students and school improvement, a learning gain could be observed between beginning of initial teacher education and two years after. However this was only limited to declarative-conceptual knowledge. Future teachers after two years of training had difficulties applying their knowledge.

Use for this review

60. The study provides further insight into the TEDS-M measurement instrument testing GPK, since it provides empirical findings from correlational and longitudinal data analysis. However, the study is limited to initial teacher education in Germany. The other test instrument shows that, besides teaching, there are other teacher tasks that could also be focused on in a standardised test.

EMW (König, Rothland, Darge, Lünemann, & Tachtsoglou, 2013)

Background Information and Summary of Study Design

61. The EMW study (*Entwicklung von berufsspezifischer Motivation und pädagogischem Wissen in der Lehrerausbildung / Change of Teaching Motivations and Acquisition of Pedagogical Knowledge during Initial Teacher Education*) is a comparative and longitudinal study conducted in the German speaking countries Germany, Austria, and Switzerland. It applies the TEDS-M test measuring GPK to the context of the Central European area. Data consists of a total sample of 6,601 pre-service teachers started teacher education in winter term 2011. They come from 31 universities/pedagogical colleges and represent a population of nearly 50,000 pre-service teachers at the beginning of their teacher education (König et al., 2013). The second occasion of measurement was conducted two years later in winter term 2013.

Test Instrument and Choices of Technical Solutions

62. See the previous section on TEDS-M for a description of the test instrument used in this study. To ensure cross-national validity of testing, pilot studies had been carried out in Austria and Switzerland supported by expert reviews (see, e.g., König & Blömeke, 2009a, b; 2012).

Major Research Questions

63. Similar to TEDS-M, the EMW study asks for the mean level of achievement of future teachers from the three participating countries in order to compare the outcomes of teacher education in Germany, Austria, and Switzerland. Also, effects of opportunities to learn and the influence of teaching motivations are to be analysed using longitudinal data.

Key Findings

64. Findings from longitudinal data analysis have not been published yet. First findings from the first occasion of measurement give insight into the relationship between pedagogical experiences future

teachers had gained *before* entering initial teacher education (e.g., through giving extra lessons to students), teaching motivations, and GPK of pre-service teachers at the very start of initial teacher education.

Use for this review

65. The study provides further insight into the TEDS-M measurement instrument testing GPK, since it provides empirical findings from another comparative sample. However, since the study is still being conducted, relevant findings from longitudinal data analysis still have to be published, limiting the use for this review at the time of writing.

SKILL (Pflanzl, Thomas & Matischek-Jauk, 2013; König & Pflanzl, under review)

Background Information and Summary of Study Design

66. The SKILL-study (*Studie zur Kompetenzentwicklung in der Lehrerinnen- und Lehrerausbildung für Berufsschulen / Study on the Competence Development during Teacher Education for Vocational Schools*) investigated 55 beginning in-service teachers at vocational schools in Austria. They were sampled out of a population of 246 teachers who had just finished their initial teacher education in Austria and who had entered the teaching profession. They were tested using the TEDS-M measurement instrument to assess their GPK. A previous study had shown that the TEDS-M test is valid for the teachers in Austria (König & Blömeke, 2009a, b). Besides, the students who were taught by the teachers tested in the study were surveyed using a questionnaire that contained scales to assess the quality of teachers.

Test Instrument and Choices of Technical Solutions

67. See the previous section on TEDS-M for a description of the test instrument used in this study. The scales assessing the teacher quality used in the student questionnaire contained items related to a teacher's instructional competence and a teacher's classroom management competence (see Figure 8 for the English translation of these scales; the German version is documented in Hopf, 2011, p. 90). They are Likert-scales with four categories ranging from "not true" (1) to "true" (4).

Figure 8: Scales assessing teacher quality used in a student questionnaire (Hopf, 2011, p. 90)

instructional competence	$\alpha = .89$
The teacher...	
- divides the lesson in sections that fit to each other.	
- explains the learning content in a comprehensible way.	
- teaches using various teaching methods.	
- knows exactly what a student is able to and is not able to.	
classroom management competence	$\alpha = .91$
The teacher...	
- clarifies which student behavior he/she expects.	
- realises everything what happens in the classroom.	
- permanently controls how students work.	
- intervenes immediately when a student starts to disrupt/misbehave.	

Major Research Questions

68. In order to analyse the significance of teacher knowledge, student scores in the teacher quality scales (teaching methods, social relationship) were aggregated on the level of the learning group and then

correlated with teacher knowledge score. The authors of the study asked for the direction (positive/negative) and the effect size of correlational analysis.

Key Findings

69. Findings show teachers' GPK correlates with both the student ratings on teaching methods competence ($r = .34, p < .01$) and the social relationship competence ($r = .36, p < .01$). That is, the higher the GPK test score, the better students evaluate the teachers' competence to apply high quality teaching methods and to foster high quality social relationships in the classroom.

Use for this review

70. The study provides further insight into the TEDS-M measurement instrument testing GPK, since it provides empirical findings on the relationship of in-service teachers' GPK and their student ratings on teacher quality in the classroom. It is crucial to see that obviously teachers' GPK as assessed by the TEDS-M instrument matters – at least from the point of view of the students taught by those teachers.

BILWISS (Kunina-Habenicht et al., 2013)

Background Information and Summary of Study Design

71. In the BILWISS project (*Bildungswissenschaftliches Wissen und der Erwerb professioneller Kompetenz in der Lehramtsausbildung* / Educational-scientific knowledge and the acquisition of professional competence during initial teacher education) future teachers' educational-scientific knowledge (“*bildungswissenschaftliches Wissen*”) is tested using a paper-pencil test instrument. The target group are future teachers at the start of the second phase (practical phase) in Northrhine-Westphalia, one out of 16 federal states in Germany. A census survey was conducted in 2011 testing the knowledge of 3,273 future teachers of all subjects and school types.

72. The authors refer to the classification by Shulman (1986), but in their study the term educational-scientific knowledge is understood as an overarching term including educational, pedagogical-psychological, and sociological knowledge that is relevant in teacher education (Kunina-Habenicht et al., 2013, p. 2).

Test Instrument and Choices of Technical Solutions

73. A completely new paper-pencil test instrument measuring educational-scientific knowledge was developed. First, a Delphi-study, in which 104 experts were asked for the most relevant topics in German teacher education programmes related to educational-scientific knowledge, was carried out. The expert group consisted of 49 persons (see for details, Kunina-Habenicht et al., 2012; Terhart et al., 2012) who were either representatives from univertarian disciplines (education, psychology, and sociology) or teacher educators from the second phase of initial teacher education that is institutionalised by the federal states (these teacher educators are in-service teachers who have been qualified to teach student teachers). Following the findings from this Delphi-study, the BILWISS project distinguishes between five topics to be tested (Kunina-Habenicht et al., 2013, p. 10):

- didactics (Unterrichtsdidaktik), containing items on lesson planning, general didactical theories, teaching methods;
- school pedagogy (Schulpädagogik), containing items on school organisation, school theories, aspects of education systems, student diversity;

- educational theory (Bildungstheorie), containing items on educational theory and history of education;
- learning/development (Lernen/Entwicklung), containing items on psychological and sociological aspects of learning and socialisation and corresponding theories, such as motivational theories;
- diagnosis/evaluation (Diagnose/Evaluation), containing items on research and assessment methods, and basic concepts of statistics.

74. Using research literature and handbooks and with the contributions from experts, 280 test items were developed.⁵ Future teachers had to recall knowledge or to apply knowledge in order to answer the test items. Multiple-choice response items as well as open-response items were included in the test.

Major Research Questions

75. At the time of writing, only the publication of Kunina-Habenicht et al. (2013) was available. There the authors report about first results from the test instrument. The scaling analysis in this article was only based on 70 items out of the 280 items included in the study.

76. The authors analyse several research questions. For the purpose of this paper, the following questions seem to be most relevant: The authors ask for the inter-correlations of the five dimensions of the test that cover the topics didactics, school pedagogy, educational theory, learning/development, diagnosis/evaluation. Moreover, they ask for mean differences by future teachers' universities.

Key Findings

77. The five subscales measuring educational-scientific knowledge in the fields of didactics (11 items), school pedagogy (16 items), educational theory (10 items), learning/development (21 items), diagnosis/evaluation (12 items) are only moderately inter-correlated. Manifest correlations range from $r = .21$ to $r = .31$. The effect size derived from an analysis of variance that specified the university where a future teacher had studied as an independent variable and the knowledge tested as a dependent variable was low ($ICC < .02$). So in general, educational-scientific knowledge seems to be a heterogeneous construct and variance in the test results can hardly be explained by the institutions where future teachers have studied.

Use for this review

78. The study aims at a broad evaluation of initial teacher education outcomes in Germany related to the specific field of educational-scientific knowledge. There is a clear intention to collect test data linked to the national curriculum particularly, whereas the authors of the study (Kunina-Habenicht et al., 2013) do not explicate whether and how their test instrument could be applied for international assessment purpose. Therefore not only the study itself, but the test instrument seems to be limited to initial teacher education in Germany. What might be interesting for this paper is the approach to conduct a broad expert review (delphi-study) to examine which topics should be chosen to measure a very broad test construct.

⁵ Items are not published in the article by Kunina-Habenicht et al., 2013, and the author has asked the Bilwiss research team for any further information, but has not yet received any more materials. Please check the project website for more information: <http://www.bilwiss.uni-frankfurt.de/index.html>

COACTIV-R (Voss, Kunter & Baumert, 2011)

Background Information and Summary of Study Design

79. The COACTIV-R project builds on a larger project, the so-called COACTIV-study (Baumert et al., 2010). Both studies were carried out in Germany. While the COACTIV-study investigated in-service mathematics teachers (see for more details Baumert et al., 2010), the COACTIV-R project aims to investigate the acquisition of teacher knowledge during the second phase of initial teacher education, i.e., the “Referendariat”, which is a practical phase following the academic study at university (for more details on the teacher education structure in Germany see König & Blömeke, 2013).

80. Among other goals of the COACTIV-R project, a new measurement instrument was developed focusing the pedagogical/psychological knowledge (PPK; see the central publication Voss, Kunter & Baumert, 2011, for more details). The definition of PPK is based on Shulman’s classification of teacher knowledge, but emphasises the psychological aspects of general pedagogical knowledge: PPK “can be defined as the knowledge needed to create and optimise teaching-learning situations across subjects, including declarative and procedural knowledge of classroom management, teaching methods, classroom assessment, and student heterogeneity” (Voss et al., 2011, p. 952).

Test Instrument and Choices of Technical Solutions

81. From a content point of view, the authors conceptualise five subdimensions to describe PPK: knowledge of classroom management, knowledge of teaching methods, knowledge of classroom assessment, knowledge of students’ learning processes, knowledge of student characteristics. Classroom management, teaching methods, and classroom assessment are assigned to the category of classroom processes, whereas learning processes and individual characteristics are assigned to the category of students’ heterogeneity (Voss et al., 2011, p. 954). Voss et al. (2011, p. 957) use 39 items in total to measure these dimensions (multiple-choice items, short-answer items, and video-based items). (Since video-based items are used to measure the subdimension of classroom management only and the scaling analysis does not differentiate according to test item modes, this instrument is described here and not placed in the following section where studies using a video-based stimulus for testing knowledge and/or skills are presented.)

82. Several pilot studies were conducted to develop and finalise the instrument including expert reviews and ratings of in-service teachers who had to rate the instrument’s content with regards to the relevance for teaching, irrespectiveness of teaching subjects, and authenticity of situations used in the test items.

Major Research Questions

83. Research questions are related to various aspects of the instrument’s validity. The internal structure was examined, assuming multidimensionality of the PPK construct. Future teachers at different stages during initial teacher education were tested, assuming differences in test performance mirroring their learning level. PPK scores were compared to other test scores (such as content knowledge). And test scores were correlated with student ratings for instructional quality of the teaching provided by the future teachers tested.

Key Findings

84. Examination of the internal structure: Both a second order factor model and a multidimensional model specifying four content subdimensions (knowledge of student’s learning processes, knowledge of student characteristics were collapsed to the broader subdimension of students’ heterogeneity) showed

good fit in structural equation modeling (confirmatory factor analysis with latent variables; Voss et al., 2011, p. 961). So, theoretical assumptions about the structure of PPK were largely confirmed by the empirical analyses.

85. Future teachers in the second year of the second phase training had higher scores compared with the future teachers in the first year of the second phase training. Mean differences were practically relevant ($d \geq .2$) for the subdimension of classroom management, whereas mean differences were only small for the other subdimensions teaching methods, classroom assessment, students' heterogeneity ($d \sim .01$; Voss et al., 2011, p. 962). Since the second phase of initial teacher education in Germany has a special focus on applying knowledge for teaching, Voss et al. (2011) interpret the mean differences as an effect of practical opportunities to learn the two groups compared have encountered to a different degree.

86. Test scores were compared with other cognitive measures: PPK correlates substantially with reasoning abilities (.58), with PCK (.42), and with CK (.24). This analysis shows that PPK is not identical with other measures of teacher knowledge or skills.

87. In a small sample of 27 future teachers, correlations were found between test scores and measures for instructional quality, which were collected via a survey of students taught by the future teachers over one school year. PPK was therefore correlated with indicators measuring cognitive activation (.42), pace of instruction (.58), student-teacher relations (.22), teachers' awareness of students' comprehension problems (.22), classroom management (.20). All correlations were positive, i.e., the higher the PPK test score, the better the instructional quality the future teachers provided (Voss et al., 2011, p. 962). These results provide first evidence that pedagogical knowledge is associated with higher quality of opportunities to learn which in turn should turn out to foster student attainment.

Use for this review

88. The COACTIV-R measurement instrument covers important subdimensions of what has been discussed in the literature related to GPK. Similar to the TEDS-M instrument, typical tasks teachers have to master are used as a starting point to develop a framework that defines the knowledge of teachers that is to be tested.

89. The project itself focuses on pre-service teachers. With a substantial sample of 746 second phase future teachers in Germany (i.e., future teachers who have already finished the first phase at university and who have entered the practical phase that is comparable with an induction phase in other countries) empirical analysis is carried out to examine the structure and the validity of the test. All in all, findings show that the instrument is of high quality (e.g., reliability, validity), so that it could serve as an example for the development of an instrument measuring GPK of teachers.

PRAXIS series from the Educational Testing Service (ETS) (ETS, 2013)

Background Information and Summary of Study Design

90. The assessment instrument "Praxis II: Principles of Learning and Teaching" (PLT) is a test highly relevant to describe and discuss in this paper. It is related to "a broad range of job-related topics" and "designed to assess a beginning teacher's knowledge" in the relevant fields (ETS, 2013, p. 2). The test measures knowledge "that is typically obtained in undergraduate courses in educational psychology, human growth and development, classroom management, instructional design and delivery techniques, evaluation and assessment, and other areas of professional preparation" (ETS, 2013, p. 2). This is regarded as professional knowledge of beginning teachers (ETS, 2013, p. 13).

91. In order to cover different grade ranges, ETS has developed four tests. They are related to early childhood, grades K-6, grades 5-9, and grades 7-12. “While the four tests cover the same topics and include materials common to all grade ranges, each test also features some material addressing the topics in ways appropriate to its particular grade range” (ETS, 2013, p. 2). The tests are aligned with US standards for teaching and teacher education such as the INTASC. In the following, the tests for grades 5-9 and 7-9 are focused on.

Test Instrument and Choices of Technical Solutions

92. There are four topic categories describing the content of the test (ETS, 2013, p. 3): students as learners (I); instructional process (II); assessment (III); professional development, leadership, and community (IV). Moreover, on a detailed level, these topic categories are differentiated again (Figure 9). The analysis of instructional scenarios (case studies) forms another category (V). Topic categories were developed on the basis of a job analysis (Robustelli & Tannenbaum, 2010).

93. The job analysis was carried out with the specific aim to develop a definition of knowledge and skills relevant for beginning teachers entering the profession, whereby that definition should form the basis for the development of the assessment instrument. First, by using standards and research literature documents, a preliminary construct was generated. That construct was reviewed and modified by experts (committee of teachers, educators from a college faculty and a department of education)⁶ resulting in 65 statements grouped into the following categories (Robustelli & Tannenbaum, 2010, p. 4): “students as learners, instructional process, assessment, and professional development, leadership, and community“. In a survey with a sample of almost 1,000 educators, these statements were judged on regarding their importance (“1 = not at all important”, “5 = extremely important”). Out of the 65 statements, 55 were being regarded as being very important (using the criterion of $M > 3.5$), and hence these statements were regarded as being relevant for the development of the assessment instrument.

Figure 9: Topics and areas covered in the ETS assessment (ETS, 2013, p. 13) as well as number of items in the Grades 5-9 Test (ETS, 2011, p. 3)

I.	II.	III.	IV.	V.
Students as Learners	Instructional Process	Assessment	Professional Development, Leadership and Community	Analysis of Instructional Scenarios
- Student Development and the Learning Process - Students as Diverse Learners - Student Motivation and the Learning Environment	- Planning Instruction - Instructional Strategies - Questioning Techniques - Communication Techniques	- Assessment and Evaluation Strategies - Assessment Tools		two case studies
21 multiple-choice items	21 multiple-choice items	14 multiple-choice items	14 multiple-choice items	4 constructed response items

94. The format of the test is as follows: It consists of two case studies (with short-answer constructed-response questions) and 70 multiple-choice response items. Test length is two hours (ETS, 2013, p. 2), and candidates are recommended to spend 70 minutes on part A (70 multiple-choice response items) and 50 minutes on part B (case studies) (ETS, 2011, p. 3).

⁶ The names and affiliations of the experts are provided by Robustelli and Tannenbaum (2010, p. 14, Appendix A).

95. As outlined by ETS (2011, p. 3), scores are reported to candidates both as a general score, as subscores related to each content area I to IV (see Figure 8), and as an additional subscore for the analysis of instructional scenarios (area V – case studies). Taking the Grades 5-9 Test as an example, subscores are measured by approximately 14 to 21 items each, whereas four items measure the performance in the text-vignettes (case studies).

Major Research Questions

96. Analysing research questions is not the primary focus of the Praxis Series Assessments. Instead, its main purpose is to meet licensing requirement (ETS, 2013). Future teachers are required to demonstrate evidence of a certain level of competence, which is defined by the Praxis assessment.

Use for this review

97. The assessment seems to be valid, but there is little information on the scaling procedure in the literature (e.g., reliability coefficients reported refer to the overall score only; see ETS, 2010, p. 60). The materials published (especially ETS, 2013) provide a good insight into the operationalisation of topics and areas covered in a variety of test items. The detailed description and alignment of topics, areas, standards, and test items constitutes a high quality example of an assessment instrument.

Studies Using Paper-Pencil-Tests to Measure a Small Segment of GPK

Introduction to the Section

98. This section is similar to the previous one, since General Pedagogical Knowledge (GPK) of pre-service or in-service teachers is measured using standardised measurement paper-pencil-test instruments. However, by contrast to the previous section, the studies outlined here focus on a small segment of GPK meaning that they measure a specific part or aspect of what could be understood by teachers' GPK.

Knowledge of Cooperative Learning (Bouas, 1996).

Background Information and Summary of Study Design

99. In this study, cooperative learning as a teaching strategy is investigated, since previous research has shown it is associated with student achievement, self-esteem, and social interaction skills (Bouas, 1996, p. 45). It seems to be essential that cooperative learning should be part of teacher education programmes, so this study's intention was "to examine the effect instruction about and participation in group work/cooperative learning in three pre-service teacher education methods classes had on future teachers' attitudes toward, knowledge about academic and social benefits related to, and pedagogical competence to organise classrooms for group work/cooperative learning" (Bouas, 1996, p. 46). Among other indicators used in the study, an assessment instrument was developed and applied to test future teachers' knowledge of cooperative learning as a teaching strategy. In the following, this will be focused on, since the test is of direct significance for this review.

Test Instrument and Choices of Technical Solutions

100. 10 multiple-choice-items to test future teachers' knowledge of cooperative learning as a teaching strategy were developed and used in the study (see Bouas, 1996, Appendix A, p. 57). Moreover, interviews were carried out and data was triangulated. The content of the test was related to a cooperative learning model. The multiple-choice-items refer to "research based academic and social benefits" that are associated with this model (Bouas, 1996, p. 53). Item examples are as follows: "Students' academic achievement suffers as a result of group work", "Cooperative learning results in students having a more positive attitude

toward school”, “Cooperative learning deters racial prejudice among students”, and the future teachers are asked to circle a “T” or an “F” if they believe the specific item is true/false or if they do not know, they are asked to circle a “DK” (for don’t know).

Major Research Questions

101. The test was applied in a pre-post-design thus allowing to measure knowledge growth of future teachers in the field of cooperative learning. Between pre- and post-measurement, future teachers were trained in cooperative learning teaching strategies.

Key Findings

102. Post means were statistically significantly higher than pre-means (Boaus, 1996, p. 53), suggesting effectiveness of the training to foster GPK in the field of cooperative learning.

Use for this review

103. From the content point of view, the measurement instrument is very focused on a special aspect of GPK. However, it is worth to account for the theoretical and evidence-based rationale to develop the test items in this study.

KADDS (Sciutto, Terjesen & Frank, 2000)

Background Information and Summary of Study Design

104. The study focuses on teachers’ knowledge of ADHD (Attention-Deficit/Hyperactivity Disorder). Presumably, teachers need to draw on this knowledge when diagnosing student behavior appropriately. Research findings show “correctly identified cases of ADHD were associated with better treatment compliance and outcome, whereas incorrectly identified cases tended to terminate treatment prematurely”, so teachers have to be trained with regards to those “characteristics of ADHD that are more effective in establishing a definitive diagnosis and more likely to promote accurate referrals.” (Sciutto et al., 2000, p. 116). 149 teachers were sampled in the New York area with an average of about 13 years of teaching experience.

Test Instrument and Choices of Technical Solutions

105. The KADDS (Knowledge of Attention Deficit Disorders Scale) is a test instrument containing 36 items. Each item consists of a statement and uses a true (T), false (F) or don’t know (DK) format (Sciutto et al., 2000, p. 117). That technical solution is chosen in order to distinguish what teachers do not know from what they believe incorrectly, for example misperceptions.

106. From a content point of view, three specific areas of knowledge of attention deficit disorders are defined (Sciutto et al., 2000, p. 117-118):

- symptoms/diagnosis of ADHD (9 items);
- the treatment of ADHD (12 items);
- general information about the nature, causes, and outcome of ADHD (15 items).

Major Research Questions

107. The study analysed psychometric indicators of the KADDS as well as frequency distributions of subscales among teachers sampled in the study. In addition, teacher background variables (e.g., age, education level) were analysed with regards to their correlation to KADDS.

Key Findings

108. The KADDS is a reliable measure ($\alpha = .81$), even for the three subscales ($\alpha = .71$ each). Inter-correlations between subscales are high ($.85 \leq r \leq .91$). This suggests “that teachers’ knowledge in one area tended to be related to their knowledge in the other areas” (Sciutto et al., 2000, p. 118). Teachers provided more correct answers on items of the symptoms/diagnosis subscale compared with the two other subscales.

Use for this review

109. Similar to the previous study (Bouhas, 1996) reviewed, this study by Sciutto et al. (2000) focuses on a specific aspect of teachers’ GPK. Again, it is worth to account for the theoretical and evidence-based rationale to develop the test items in this study. However, critically reviewed, the question arises whether the response format (true, false, don’t know) is appropriate to purely measure teacher knowledge. Because there is a substantial correlation between KADDS and teachers’ confidence in their ability to effectively teach an ADHD child ($r = .29$; see in detail, Sciutto et al., 2000, p. 120) one might assume that the KADDS is a conglomerate of teacher knowledge and beliefs about ADHD. To conclude, this study is worth to account for when looking at initiatives to assess teachers GPK.

Studies Using a Video-based Stimulus for Testing Knowledge and/or Skills*Introduction to the Section*

110. Similar to the previous sections, studies outlined here measure General Pedagogical Knowledge (GPK) of pre-service or in-service teachers using standardised measurement paper-pencil-test instruments. However, by contrast to studies mentioned in the previous sections, the studies outlined here use videos as a stimulus in the item stem, an assessment format which is frequently referred to as “video-vignette” or “video-cued testing“. Video-based assessment instruments are used to address the contextual nature of the classroom situation. They are considered to improve the measurement of teacher knowledge when compared with the classical paper-pencil-test.

TEDS-Follow Up (König et al., 2014)*Background Information and Summary of Study Design*

111. In 2008 IEA’s Teacher Education and Development Study – Learning to Teach Mathematics (TEDS-M) was carried out (see previous section for further details). In Germany as one of the 17 participating countries, a follow-up has been initiated (TEDS-FU; Blömeke, Kaiser, & König, 2009). TEDS-FU samples German in-service teachers of mathematics who had participated in TEDS-M and collects data on their knowledge via TEDS-M paper-and-pencil test instruments. In both studies TEDS-M and TEDS-FU the TEDS-M paper-pencil-test instrument was used to measure GPK (König et al., 2011). Moreover, TEDS-FU makes use of video-vignettes providing typical classroom situations as a stimulus followed by various test items including items on generic pedagogical challenges such as classroom management (König et al., 2014).

Test Instrument and Choices of Technical Solutions

112. See the previous section on TEDS-M for a description of the test instrument measuring GPK. Besides the GPK test derived from TEDS-M, a video-based assessment was developed in TEDS-FU. In TEDS-FU, three 3- to 4-minutes video-vignettes were developed and provided to early career teachers of mathematics. After each video-clip, the early career teachers of mathematics had to respond to several test items. Some of them were related to mathematical pedagogical content knowledge (MPCK), others to general pedagogical knowledge (GPK). In the following we focus on those items related to GPK.


113. A theoretical framework was developed focusing on the skills to notice and to interpret classroom situations (see for details, König et al., 2014). Both skills are relevant for the successful mastering of teaching, and various frameworks have already been developed to underline this (König et al., 2014, p. 78): “To summarize, these frameworks have in common to distinguish between mere observing the concrete situation provided by a video-clip, whereby meaningful aspects are identified (“noticing”), and more elaborated cognitive processes such as interpreting or analyzing the situation (“interpreting”). On a detailed level of differentiating cognitive processes, it might not always be possible to sharply discriminate between noticing and interpreting, since presumably these cognitive processes are interdependent or might even be localized on a continuum. So in TEDS-FU, the basic distinction between noticing and interpreting was applied as a heuristic when video-vignettes were developed.”⁷

114. Items measuring the cognitive demand of noticing were multiple-choice items with a Likert-rating scale (four categories) ranging from “completely agree” to “completely disagree”. About half of them were related to the precision of teacher perception (e.g., “The teacher presents the lesson’s task visually AND acoustically”), the other half measured teacher perception holistically (e.g., “Most students take an active part in the lesson”). The teachers had to mark the extent of their agreement. Each statement referred to the teaching in one video-clip specifically. An expert rating was carried out to define which category of each item’s rating scale was to be scored as “correct” (see for further details, König et al., 2014).

115. Items measuring the skill to interpret general pedagogical classroom situations were open response items that would allow teachers to provide cognitively more complex statements. Figure 10 shows an item example that is related to one of the video-clips resulting from a mathematics lesson on computing the volume of a box. During that lesson, students had to work in pairs, whereby three pairs and their way of cooperation were focused on more closely in the video-clip. The item example shown in Figure 10 now asks the teachers for an in-depth analysis of the three student pairs’ cooperation from a pedagogical perspective. To help teachers recalling the student pairs, three screen shots of the student pairs were provided in the item stem (see Figure 10).

⁷ Thus “noticing” is more used as a synonym for “observing” here. A detailed overview of how terms such as noticing, observing etc. are currently used in the literature is provided by Kaiser et al. (under review).

Figure 10: Item example for the pedagogical skill to interpret a classroom situation (König et al., 2014, p. 82)



In the video, three student pairs' cooperation was focused on more closely. Consider these cooperation processes from a pedagogical perspective.

Describe in a contrasting mode for each pair the relevant aspects how they have carried out their cooperation.

Major Research Questions

116. Beginning mathematics teachers' data from the GPK test and the video-based assessment were analysed with respect to several research questions. One asks whether the current GPK of early career mathematics teachers is inter-correlated with their skills to notice and interpret classroom situations, which are tested via video-vignettes. Also, the study examines whether early career mathematics teachers' performance to notice and interpret classroom situations can be predicted by their GPK acquired during initial teacher education.

Key Findings

117. Teachers' skills to notice and interpret differ. Interpreting correlates with the current level of GPK, whereas noticing does not. GPK at the end of teacher education neither predicts noticing nor interpreting, which suggests teachers' cognitions are reorganised during the transition into teaching.

118. Moreover, there is a large difference in the GPK between the time point end of initial teacher training and four years later when teaching. Teachers obviously acquire GPK even after having finished initial teacher education.

Use for this review

119. The TEDS-FU study first gives insight into the test design of a video-based assessment of general pedagogical skills to notice and interpret typical classroom situations from a pedagogical perspective. The skill to notice and the skill to interpret are of different quality. This shows that what a teacher knows and is able to do is of multidimensional nature, a finding that should be accounted for by any assessment instrument measuring teacher knowledge.

120. Besides, the study gives further insight into the TEDS-M test measuring GPK via paper-pencil. In-service teachers most likely improve their GPK through experience, professional development, and informal opportunities to learn (see, for more details, König, Blömeke, & Kaiser, under review). Also, the TEDS-M test seems to be an important instrument not only to investigate the effectiveness of teacher education, but also to measure the in-service teacher's knowledge base.

121. Although TEDS-FU does not link teacher knowledge and skills to student achievement data, it approaches the measurement of teacher knowledge with an instrument that most likely is closer to performance in the classroom than just paper-pencil assessment focusing on declarative knowledge. For the purpose of this review this distinction seems to be important.

CME (König & Lebens, 2012a)

Background Information and Summary of Study Design

122. The study focuses on teachers' classroom management expertise (CME). CME is regarded as an essential component of professional teacher competence. The measurement of CME represents a critical issue for research and practice at the intersection of psychological and educational science. To account for the significance of CME measurement, the study proposes an instrument framed within theoretical and methodological considerations. A major current focus is the implementation of video-vignettes to provide a more ecologically valid measurement of pedagogical knowledge and proficiency required for the successful mastering of classroom management challenges.

Test Instrument and Choices of Technical Solutions

123. The study provides a methodological and theoretical framework for the video-based measurement related to three cognitive dimensions of CME: accuracy of perception, holistic perception and justification of action (see also König et al., 2014, p. 78-79):

- *Accuracy of perception* is relevant, since research on teacher expertise has shown expert teachers identify relevant instructional situations seen in a video-vignette assessment more precisely and correctly than do novices (e.g., Sabers, Cushing, & Berliner, 1991).
- *Holistic perception* (Bromme, 2001) is important, since expert teachers reconstruct and anticipate the context of instruction and engage in reflecting alternative problem-solving strategies, whereas novice teachers notice classroom situations step by step due to the heterogeneous structure of their knowledge (Bromme, 1992).
- *Justification of action* is most important when classroom situations are interpreted or analysed (e.g., "Why is does the teacher stand in front of the board and what is the reason behind the specific student grouping visible on the screen?"). The interpretation of events goes beyond generating mental representation (as covered by the holistic perception), as it strongly depends on reframing and transforming knowledge.

124. In the research article (König & Lebens, 2012a), one video-vignette is described in detail providing information on the item set related to the video-clip used for measurement. In a follow-up, a thorough research instrument consisting of four such video-vignettes has been developed (König & Lebens, 2012b; König & Rothland, under review).

Major Research Questions

125. An expert-novice paradigm was adopted to investigate whether a paper-pencil test containing questions about a briefly presented video-vignette can adequately discriminate between a sample (n= 108) of experienced teachers and pre-service teachers of the first and second phase of initial teacher education in Germany.

Key Findings

126. Results indicate that the proposed instrument can reliably detect expertise-related differences in relevant knowledge structures. Overall, the instrument produces valuable insights into the cognitive dimensions of CME in experienced and novice teachers. The outcome of the study carries implications for CME measurement and can be potentially applied to improve CME development in teacher training. The specification of the cognitive correlates and determinants of CME represent a crucial issue for future research.

Use for this review

127. The study shows how the video-based assessment approach can be conducted and it provides important insight into a quantitative approach to novice-expert-comparisons, since the novice-expert research paradigm has largely been applied in qualitative studies (see below the section on phenomenological studies describing teacher knowledge).

OBSERVE (Seidel, Blomberg & Stürmer, 2010)*Background Information and Summary of Study Design*

128. Seidel et al. (2010) have developed an assessment instrument called OBSERVER. It consists of video clips of classroom situations used as item prompts. Together with item statements related to these prompts, a teacher's ability to apply pedagogical knowledge is tested.

129. The instrument's framework contains teaching and learning components: goal clarity and coherence; teacher support; learning climate; experiments and scientific inquiry. Furthermore, observation requires the following cognitive processes: to identify and describe relevant components of teaching and learning processes (description); to explain an observed situation by applying scientific theories (explanation); to predict possible effects of a given situation with regard to future teaching and learning processes (prediction).

Test Instrument and Choices of Technical Solutions

130. Video clips from authentic classroom teaching are used. They have to represent the framework's teaching and learning components. Test length is aligned to 90 minutes testing time which allows for six video clips to be included and relevant test items. Example items are: "The teacher clarifies the tasks that the students have to accomplish during the lesson" (describing; component goal clarity), "The students have the possibility to experience autonomy in their learning processes" (explaining; component learning climate), "The students are able to experience themselves as self-regulated learners" (predicting, component teacher support).

Major Research Questions

131. As a first step, the assessment instrument was evaluated by student teachers and in-service teachers regarding its usability (Seidel et al., 2009; 2010).

Key Findings

132. Overall acceptance was good. Pre-service teachers as well as in-service teachers evaluated the classrooms situations shown in the video-clips as being a helpful tool to analyse lessons, to discuss teaching and learning. Video-clips were regarded as being meaningful, authentic, and interesting (see for details, Seidel et al., 2009).

Use for this review

133. Regarding the use of video-vignettes, the approach of the OBSERVE-instrument is a good example. However, there is the critical question whether an assessment using only multiple-choice-items is appropriate to complex issues of testing applied knowledge. This will be discussed in Part 3 in more detail.

TSC – Thinking in Science Classroom (Zohar, 2004)

Background Information and Summary of Study Design

134. The study by Zohar (2004) is situated in the context of the project *Thinking in Science Classroom* (TSC). It is a study with 83 Israeli junior and high school science teachers, investigating “elements of pedagogical knowledge when students’ higher order thinking was an explicit and focused instructional goal.” (Zohar, 2004, p. 293).

Instrument and Choices of Technical Solutions

135. Teachers watched a video “showing a student who is struggling to solve a computerized reasoning task” (Zohar, 2004, p. 297) and they were asked several questions when the video was stopped. The one analysed by Zohar (2004, p. 298) is as follows: “Suppose Anna [the student in the video] is your student. What you just saw in the video had happened in your own classroom. How would you react? What feedback would you give her?”

136. Teachers had to write down their responses. Responses were analysed using coding rubrics that formed qualitative categories. Then the number of responses in each category was counted for frequency analysis (Zohar, 2004, p. 298, p. 300). The instrument is exploring what teachers would respond rather than a measurement instrument.

Major Research Questions

137. Zohar (2004) does not explicitly formulate research questions, but establishes data-driven analysis categories.

Key Findings

138. Teacher responses were classified into four categories (Zohar, 2004, p. 300), which were collapsed into two broader categories (a and b, c and d):

- “(a) explicit guidance that gives detailed instructions regarding correct investigation procedures (e.g., “You must control variables. You should change only one factor in each experiment.” or “Do another experiment. Change only the variable ‘light’ and leave everything else as it was before.”);
- (b) guidance through a series of directive questions (e.g., “I would ask her a series of questions, directing her to change only one variable at a time.”);

- (c) guidance through a series of questions that lead to conflict (e.g., “I would ask her a series of questions that would make her realize that she got contradictory results.”); and
- (d) a choice not to intervene at that stage of the learning process (e.g., “At that stage, I would not say anything, and I would not guide her. It’s better to let her experiment and struggle on her own, in a free way, because I think this is one of the goals of the activity.”).

Use for this review

139. In this study, it does not become clear how knowledge is investigated. Instead, it gives the impression that there is no clear definition of GPK, since knowledge and beliefs are mixed up (see description on p. 300). It is a very explorative study.

Studies Rating the Demonstration of Teaching Skills

Introduction to the Section

140. When trying to go beyond paper-and-pencil testing or using video-based stimuli, we can identify studies that intend to rate demonstrations of teaching skills. In the following, we focus on studies that, for example, analyse lesson plans as an approach of competence measurement.

PACT (Pecheone & Chung, 2006; www.pacttpa.org)

Background Information and Summary of Study Design

141. *Performance Assessment for California Teachers (PACT)* “was developed in response to a California State mandate (SB 2042), requiring teacher preparation programs to use performance assessments as one measure in making credentialing decisions.” (Pecheone & Chung, 2006, p. 22). The PACT assessment was pilot tested over five years beginning in 2002 about its validity for detecting competence of the individual (future) teacher. In 2007 a technical report summarising validity and reliability studies of the model was published (Pecheone & Chung, 2007).

142. PACT aims at assessing future teachers’ *performance* and explicitly does not intend to test teacher (pedagogical) knowledge (Pecheone & Chung, 2006, p. 33). PACT has been applied as a measure of teacher quality for licensure. It was approved by the *California Commission on Teacher Credentialing (CCTC)* (Sandholtz & Shea, 2012).

Test Instrument and Choices of Technical Solutions

143. PACT does not contain a test measuring teachers’ general pedagogical knowledge. Instead, it collects and uses different kind of data. Its design is organised around a “Teaching Event” (TE), for which a pre-service teacher is required to complete five components that are related to one of his or her units of instruction. These components are (PACT Consortium, 2004, pp. 9):

1. a description of their teaching context,
2. a planning overview and rationale for a 3-5 lesson learning segment with a central focus,
3. one or two videotapes of instruction from these lessons accompanied by commentary describing the instruction that took place each day and in these excerpts,

4. an assessment plan and analysis of samples of student work from one assessment given during the learning segment, and
5. written reflections on instruction and student learning.

144. Because pre-service teachers are required to plan, to instruct, to assess, and to reflect, they presumably make use of general pedagogical knowledge for teaching. However, since the “Teaching Events” are subject-specific (see Pecheone & Chung, 2006, for details), the PACT does not provide an isolated measure related to general pedagogical knowledge or skills. By contrast, “the Teaching Events are designed to measure and promote candidates’ abilities to integrate their knowledge of content, students, and instructional context in making instructional decisions and reflecting on practice.” (Pecheone & Chung, 2007, p. 5).

145. After data collection, the “Teaching Events” are scored along the following tasks: planning, instruction, assessment, reflection, and academic language. So called “guiding questions” (GQs) with coding schemes are used to lead the scoring. For this, a 4-point continuum is used in order to rate each pre-service teacher’s performance (Pecheone & Chung, 2006, p. 25): (1) standard not met on the item, (2) standard met on the minimum level, (3) advanced level of performance, (4) superior level of performance. For example, for the task “planning”, the following guiding questions were used by the raters who had to score elementary literacy teaching event (Pecheone & Chung, 2006, p. 27):

- “Access to curriculum – How does the instructional design make the curriculum accessible to the students in the class?”
- Coherent instructional design – How does the instructional design reflect a coherent approach to the literacy curriculum?
- Balanced instructional design – How does the instructional design reflect a balanced approach to the literacy curriculum?
- Student needs and characteristics – How does the instructional design reflect and address student interests and needs?
- Assessment alignment – How well are the learning goals, instruction, and assessments aligned?”

Academic language is evaluated by the following item (Pecheone & Chung, 2006, p. 27):

- “How does the candidate’s planning, instruction, and assessment support academic language development?”

Major Research Questions

146. When PACT was pilot tested, research questions were related to pre-service teachers’ performance, but also to the validity of the measurement instrument.

Key Findings

147. Key findings are summarised by Pecheone and Chung (2007, p. 5), such as the following:

- Irrespective of the subject, pre-service teachers showed higher scores on the tasks “planning” and “instruction” when compared with the “assessment” and “reflection” tasks. Further, they did not show high performance on the task “academic language”.
- “Teacher educators who participated in the development and design of the assessments were asked to judge the extent to which the content of the Teaching Events was an authentic representation of important dimensions of teaching. Another study examined the alignment of the TE tasks to the California Teaching Performance Expectations (TPEs). Overall, the findings across all content validity activities suggest a strong linkage between the TPE standards, the TE tasks and the skills and abilities that are needed for safe and competent professional practice.”

Use for this review

148. Although PACT is subject-specific and does not measure knowledge on a scale, it provides important insights into a teacher performance assessment that is very close to typical tasks teachers have to master. From that, issues related to teachers’ GPK may be reconstructed. The assessment seems to be valid, but there is little information on the scaling procedure in the literature.

PLANVOLL (König, Buchholtz & Dohmen, under review)*Background Information and Summary of Study Design*

149. The PLANVOLL study analyses written lesson plans designed and applied by future teachers who were in their second phase of initial teacher education in Germany. The second phase is the practical part of initial teacher education in Germany, following the academic study at university thus comparable with the induction phase of teacher education systems in other countries (for more details on the teacher education structure in Germany see König & Blömeke, 2013). During this 1.5 to 2 year phase (length varies across federal states in Germany), future teachers have to submit several times written lesson plans to their teacher educators who then attend, observe, and assess the lesson taught by the future teacher. The PLANVOLL study collected the very first and the very last of such lesson plans of 106 future teachers during their second phase allowing longitudinal data analysis. The sample was recruited in the federal state of Berlin, where future teachers had to run through the second phase with a length of two years.

150. The PLANVOLL study aimed at the development of a measurement instrument that would allow quantifying future teachers’ decision making when planning a lesson for their specific learning group. Thus the assessment instrument is not a paper-pencil test, but it consists of 11 indicators applied to examine whether a future teacher has accounted for specific decisions in his or her lesson plan. This approach aims at an examination of teachers’ decision making related to the learning group they actually teach (not a hypothetical one). This is regarded as a typical task teachers have to master, and a challenge in which expert teachers outperform novice teachers, as known from phenomenological research on teacher expertise. The study concludes from its analysis results to future teachers’ general pedagogical knowledge, i.e., situational and conceptual knowledge related to lesson planning.

Test Instrument and Choices of Technical Solutions

151. The assessment instrument uses 11 indicators that are applied to examine the quality of written lesson plans. The main focus of the instrument is a specific task every teacher has to master during lesson

planning, irrespective of the subject, the school type, or the educational level of the learning group: to align the tasks students have to work on during the lesson to the students' pre-knowledge. This is operationalised by analysing descriptions of the learning group (e.g., how heterogeneous they are concerning their pre-knowledge, how their pre-knowledge is described) and the task (e.g., whether the task is appropriate to the students' pre-knowledge or whether a variation of tasks is provided to students according to their diverse learning needs) and by analysing connections that are done between those elements (e.g., how the tasks are aligned to lesson goals, student grouping, evaluation and feedback etc.). The indicators are dichotomous and are applied by two raters working independently. Cohen's Kappa is used to determine inter-rater reliability (.83) showing coding the lesson plans with the developed scoring schemes was successful. The 11 items are examined by using Item-Response scaling methods showing a reliable measure of lesson planning competence (EAP-reliability = .70).

Major Research Questions

152. Research questions of the PLANVOLL study refer to future teachers' acquisition of general pedagogical knowledge and skills to plan a lesson where student tasks are aligned to students' pre-knowledge. Following the research on teacher expertise, the study assumes that in the course of their training, future teachers' scores increase from t_1 to t_2 , but that especially solution probability of those items measuring connections between planning elements increase compared with the descriptive items.

Key Findings

153. After about 1.5 years of training, future teachers' competence to plan a lesson increases significantly ($d = .8$). Items connecting planning elements are more important for this learning gain than items measuring the description of isolated planning elements. Correlation analysis with GPA and teacher education grades as well as pedagogical beliefs provide evidence for convergent and discriminant validity of the competence measure.

Use for this review

154. The PLANVOLL study provides important insight into a measurement approach that is different from usual paper-pencil testing. Because real lesson plans of the teachers sampled are analysed, the approach used in the PLANVOLL study is similar to parts of the assessment conducted in PACT. However, the PLANVOLL study focuses on a generic task teachers have to master when planning a lesson, so the measurement allows to reconstruct general pedagogical knowledge of teachers.

Phenomenological Studies Describing Teacher Knowledge

Introduction to the Section

155. There are a substantial number of qualitative studies describing teacher knowledge. Since it is not the main focus of this paper to review such studies extensively, two criteria are used to select studies that will be summarised or mentioned in this section. First, studies describing differences between experts and novices are summarised with regards to the question of how they are related to general pedagogical knowledge. Second, qualitative studies are reported that describe general pedagogical knowledge explicitly. Therefore, studies that have a stronger focus on pedagogical content knowledge (e.g., Tan et al., 2011) are not included into this section.

*A Study on GPK related to Key Pedagogical Principles (Atjonen et al., 2011)**Background Information*

156. A group of Finnish schools were randomly sampled, and a total of 362 fourth grade and 1989 seventh to ninth grade teachers participated in a survey using a questionnaire. The sample represents the population of Finnish comprehensive school teachers (Atjonen et al., 2011, p. 279).

Instrument

157. In the survey questionnaire, Likert-scale items and open-ended questions were used. Main pedagogical principles were investigated using three open-ended questions (Atjonen et al., 2011, p. 278):

“B1. What is the pedagogical principle (or principles), which you consider to be the most important for your own educational and instructional work as a teacher?

B2. If you have not been able to follow the most important pedagogical principles, what were the most important obstacles that hindered you?

B3. Describe a teaching-learning event or period, which was exceptionally successful in your opinion. Include information on objectives, pupils, teaching methods, or any other important factors.”

158. A category system was developed to code conceptions that were described or mentioned in the teacher’s response. Atjonen et al. (2011, p. 280) use five major categories to summarise pedagogical principles uttered by teachers when they responded to question B1:

- teaching-learning arrangements and methods;
- relationship to pupils and learning environment;
- general educational principles;
- expected skills and attitudes;
- other principles.

Similarly, they used three main categories to code responses on B2 (Atjonen et al., 2011, p. 282):

- obstacles related to the requirement of learning;
- obstacles related to pupils and groups of pupils;
- obstacles related to teachers, pupils’ background or other issues.

And finally they used three categories to group successful pedagogical events uttered in responses to question B3 (Atjonen et al., 2011, p. 285):

- relationship to pupils and learning environment;
- teaching-learning arrangements and methods;
- other principles.

Major Research Questions

159. Atjonen et al. (2011, p. 277) ask for frequency distribution of the coded principles and differences by gender, class size, school level, school subject.

Key Findings

160. This study is primarily interesting with regards to the question related to the content of pedagogical knowledge, here specifically key pedagogical principles, how they can be grouped etc.

Use for this review

161. These findings may contribute to discuss the question, which content areas to be covered and types of knowledge to be assessed.

A Study on GPK related to Classroom Management (Garrahy et al., 2005)

Background Information

162. A convenience sample of 20 physical education teachers was used. Interviews were conducted.

Major Research Questions

163. What do the teachers know about management (Knowledge Content)? How was that knowledge gained (Knowledge Origin and Influence)? Did the knowledge change over time (Knowledge Evolution)?

Key Findings

164. One should keep in mind that although researchers wanted to investigate knowledge change over time, they did not carry out a longitudinal study. So it seems this study is primarily interesting with regards to the first question related to the content of pedagogical knowledge. With regards to knowledge content, it seems to be relevant to list the themes according to which teachers verbalised their knowledge base (Garrahy et al., 2005, p. 60):

- the need for consistency in teachers' interactions with students and in establishing routines; the need for knowledge that involved a humanistic approach (speaking with children individually, developing mutual respect, modeling desired behaviors, knowing the students);
- the need to model the behaviors desired and to acknowledge behavior when demonstrated by students;
- the need for strategies for teaching and reinforcing teacher expectations.

Use for this review

165. These findings may contribute to discuss the question, which content areas are to be covered and which types of knowledge are to be assessed.

Studies describing differences between experts and novices

Summarised Information (following Hogan et al., 2003)

166. A number of studies have been conducted to investigate novice-expert-differences with regard to issues of general pedagogical skills and knowledge. Although they do not test knowledge, they provide valuable insights into issues such as (cf. Hogan et al., 2003):

- classroom management strategies (Clarridge & Berliner, 1991; Sabers, Cushing, & Berliner, 1991; Swanson et al., 1990);

- effective communication techniques (Sánchez, Rosales, & Canedo, 1999) ;
- abilities to recall and reflect on classroom events (Allen & Casbergue, 1997; Gonzales & Carter, 1996);
- the ability to foster meaningful classroom environments (O'Connor & Fish, 1998).

Methods used

167. Think aloud protocols (e.g., Swanson et al., 1990), viewing a teaching lesson (e.g., Sabers et al., 1991), recalling teaching events (e.g., Gonzales & Carter, 1996), structured interview (e.g., Allen & Casbergue, 1997).

Key Findings

168. The studies provided insights into qualitative differences of cognitive representations between experts and novices related to general pedagogical issues such as:

- Expert teachers identify and perceive more meaningful events compared than novices (e.g., student misbehavior). Novices have a narrowed view on what they have seen (Clarridge & Berliner, 1991).
- Experts link what they see with strategies they know (e.g., how to deal with student misbehavior), novices do not know how they would deal with it or they even do not recall the relevant event (e.g., Sabers et al., 1991).
- Experts have better skills to communicate with students (Sanchez et al., 1999), e.g., to activate their prior knowledge at the beginning of the lesson or to present new information in an organised way. They use different strategies to communicate with their students (O'Connor & Fish, 1998).

PART 3: DISCUSSION OF CONCEPTUAL ISSUES

General Pedagogical Knowledge vs. Pedagogical Content Knowledge vs. Content Knowledge

Introduction to the Section

169. In order to structure relevant cognitive domains of teachers, mainly three knowledge areas are distinguished (cf. Shulman, 1986; Bromme, 1997; Baumert & Kunter, 2006): Content Knowledge (CK), Pedagogical Content Knowledge (PCK), and General Pedagogical Knowledge (GPK). The growing body of research on the knowledge of teachers and future teachers has a special focus on the knowledge related to mathematics (Krauss et al., 2008; Hill et al., 2005; Schmidt et al., 2007; Baumert et al., 2010; Tatto et al., 2012; Kleickmann et al., 2013). So the studies mainly focus on CK and PCK of teachers and future teachers, and the knowledge is exemplified by the discipline of mathematics.

170. As described in Part 2, our methodological approach was to de-select studies that did not measure GPK, but related their measurement to PCK. Now having reviewed studies that aim at a measurement of GPK in the closer sense, the question has to be discussed anew, how GPK differs from PCK and, especially, how GPK can be empirically separated from PCK.

How to separate GPK from subject-related Teacher Knowledge

171. In general, it seems to be a specific research question how GPK can be separated from PCK. Whereas theoretical assumptions have been established and discussed (e.g., Shulman, 1986; 1987; Grossman, 1990), empirical findings on the question how GPK can be or has to be separated from subject-related teacher knowledge, especially teachers' PCK, are very rare. This is due to the small number of studies that investigated both the PCK and the GPK of teachers, for instance, to carry out correlational analysis on construct validity. In the studies reviewed, there seem to be two studies only that provide evidence on this.

Correlational Analysis in TEDS-M on GPK vs. PCK/CK

172. In TEDS-M, paper-pencil tests to measure future teachers' CK and PCK in mathematics were developed and applied (see, in detail, Tatto et al., 2008; 2012). Regarding PCK, topics of the test include pre-active curricular and planning knowledge as well as interactive knowledge about how to enact mathematics for teaching and learning. The items were related to areas such as establishing learning goals, knowing different assessment formats or linking teaching methods, or diagnosing typical student responses, including misconceptions, and providing appropriate feedback. Regarding CK, topics of the test cover core areas of school mathematics in middle school, namely numbers, algebra, geometry, and stochastic/data and chance.

173. IRT scaling analysis was applied to examine the multidimensional teacher knowledge structure. Each of the three knowledge domains could be measured as a reliable test score. Then, inter-correlations of the three scores were examined. Regarding the secondary level mathematics teacher samples from the US, Taiwan, and Germany, GPK was positively correlated with PCK and CK. Interestingly, the size of the correlations differed across countries (USA: .16/.16, Taiwan: .14/.11, Germany: .30/.29) but not between GPK and PCK vs. GPK and CK (see for further details, Blömeke & König, 2010, p. 275). Variation in the size of correlations was explained by country differences to the extent to which they stratify their school and teacher education system. Since Germany has closely aligned teacher career paths to its tripartite school system, it is most likely that a correlation of about .30 in contrast to other country's correlations

between .11 and .16 goes back to differences in opportunities to learn for the various teacher types (see for details, Blömeke & König, 2010).

Correlational Analysis in COACTIV-R on GPK vs. PCK

174. Also in the COACTIV and the COACTIV-R study that were carried out in Germany, paper-pencil tests to measure German in-service (COACTIV) and German pre-service (COACTIV-R) teachers' mathematical content knowledge (CK) and mathematical pedagogical content knowledge (PCK) were developed and applied (see for details on these two tests, Baumert et al., 2010). Voss et al. (2011, p. 962) report on positive and statistically significant ($p < .05$) correlations between future teachers' pedagogical knowledge and their pedagogical content knowledge (.42) as well as their content knowledge (.24).

Use for this review

175. To sum up, COACTIV-R, but especially the international data from TEDS-M show correlations are not high in any country, so one can assume GPK and PCK do form different constructs in various cultural contexts and teacher education systems, which legitimates to investigate GPK for analytical reasons without directly accounting for an additional measurement of PCK. However, taking into consideration teacher competence is multidimensional, one should be aware that a sole focus on the measurement of GPK leaves behind an empty field related to subject matter and its domain-specific pedagogy (i.e., pedagogical content knowledge). This is especially true when looking how teacher knowledge is related to student achievement. Here one can assume that PCK and GPK have an impact on overall teaching performance, and PCK and GPK are more important than CK (Darling-Hammond, 2000).

Definition of Content

Intro to the Section

176. On the international, but – at least seen from a German perspective – also on the national level it is a great challenge to determine what can be understood by the term GPK and what can be incorporated into this knowledge domain. There are several reasons for this. At least, this is due to the content diversity of general pedagogy as it is implemented as an opportunity to learn in teacher education and the lack of empirical studies on the GPK of (future) teachers.

177. It might be obvious that the GPK includes principles and theories of instruction and learning that are not subject-matter bound, including classroom management. In addition to that, knowledge about counseling and nurturing students' social and moral development (“Erziehung”) as well as the knowledge about school development (“Schulentwicklung”) might also be included into the area of general pedagogy (cf., e.g., for Germany, KMK, 2004).

178. So, for example, typical course content in general pedagogy contain (see KMK, 2004; Kunina-Habenicht et al., 2012): teaching (including theories about teaching, teaching strategies, principles, and methods, lesson planning), nurturing students (including disruptive behavior in the classroom, values, moral aspects of teaching), assessment (including educational psychology research), schooling (historical, philosophical, sociological aspects), professionalisation of teachers (including teacher role, vocational requirements).

179. Since there is a lack of empirical studies investigating the GPK, many key questions remain unanswered. A decade ago, there were virtually no theoretical concepts or empirical studies showing how to conceptualise the GPK or how it could be structured into dimensions and categories (cf., e.g., Baumert & Kunter, 2006). Very recently, this has started to change. In Part 2 of this review, some studies are reviewed that deal with principle questions of how GPK can be theoretically and empirically

conceptualised, measured, and validated. A typical challenge that those studies have to face is related to the following question: How can content of GPK be defined and operationalised in test items?

Content of GPK and its Operationalisation

180. Defining what GPK actually is and what it is not, which content to be included into an assessment is very difficult, and to identify the content of teacher GPK is a challenging task. As phenomenological studies show (e.g., Garrahy et al., 2005, p. 61), “research is needed to continue to clarify the exact content of teachers’ knowledge”.

Defining Content from a Classification of Typical Vocational Tasks Teachers have to Master

181. To conclude from what has been said about the studies in the previous part, one possible approach is to describe, observe, and analyse what typical challenges are in the classroom that teachers have to master and then to characterise the knowledge teachers need for mastering such challenges. This approach makes use of what we already know about typical challenges in the classroom, especially solid findings from process-product research.

182. The research related to general pedagogy contains examination of teacher classroom management strategies, the organisation of student learning and assessment, and effective communication techniques and teaching strategies/methods. Teachers need to recall and to reflect on classroom events and they should be able to foster meaningful classroom environments. They also have to know educational goals, values, and purposes. All this is a knowledge base mainly irrespective of the subject (see also Hogan, Rabinowitz & Craven, 2003). Then, didactics is also seen as a relevant category (Kansanen, 2002). Reflecting on the studies reviewed, general didactic knowledge can be a relevant part of assessing teacher general pedagogical knowledge (cf. König et al., 2011; Kunina-Habenicht et al., 2013), even when compared internationally (König et al., 2011; 2014).

Defining Content from Formal Disciplines / Academic Underpinnings for General Pedagogy

183. Another approach seems to be that formal disciplines could be named providing the intellectual underpinnings for the field of education such as educational psychology, sociology of education or history of education. For example, in MT21, which served as a kind of pre-study for TEDS-M and the GPK measurement instrument, Blömeke et al. (2008, p. 751) used the three academic disciplines didactics, educational psychology, and educational sociology to define “three areas of general pedagogical knowledge: lesson planning, assessment and dealing with heterogeneity, taking socio-economic differences as an example.” Also in the BILWISS-study, using the method of a Delphi-study, content areas were specified with regards to certain academic fields.

A summary of paper-pencil tests to measure GPK extensively

184. The tests presented in the first group in Part 2 (i.e., studies using paper-pencil tests to measure GPK extensively) show similarities and differences (see Figure 11). The TEDS-M test and the COACTIV-R test show similarities in their focus on instruction/classroom management and their differentiating subdimensions such as classroom management, assessment. By contrast, the PLT covers instruction and assessment as well, and, similar to the COACTIV-R test, the PLT is related thematically to students as learners. Moreover, the PLT covers content areas communication techniques, which is close to what the TEDS-M test and the COACTIV-R test partly cover with the subdimension of classroom management, but it also covers an area that is more distant from teaching, namely teacher professionalism which deals with the profession and the community.

Figure 11: Topics and areas covered in paper-and-pencil-tests measuring GPK extensively

	TEDS-M	LEK	BILWISS	COACTIV-R	PLT
TEACHING	structure		didactics	teaching methods	instructional process
	classroom management			classroom management	
	motivation adaptivity		learning/ development	students' heterogeneity (student characteristics, students' learning process)	students as learners
	assessment		diagnosis/evaluation		assessment
NON-TEACHING TASKS		school improvement	school pedagogy		professional development, leadership and community
		nurturing and educating students	educational theory		

185. To sum up, when we look at GPK for teaching, three broad fields can be identified when looking at all these tests (TEDS-M, LEK, BILWISS, COACTIV-R, PLT) reviewed:

- *instructional process* (including teaching methods, didactics, structuring a lesson, and classroom management),
- *student learning* (including their cognitive, motivational, emotional individual dispositions; their learning processes and development; their learning as a group taking therefore into account student heterogeneity and adaptive teaching strategies), and
- *assessment* (including diagnosing principles irrespective of the subject, evaluation procedures)

186. By contrast, non-teaching tasks seem to be heterogeneous leaving the question of how to identify consensus about such broader fields open to future research.

How the other tests reviewed fit into this summary

187. As a second step of analysing the summary previously presented, the question arises how all other studies reviewed in Part 2 can be assigned to the three broader field of GPK content. As the proposed assignment in Figure 12 shows, all assessment instruments reviewed can be sufficiently assigned to the three broader fields:

- the studies using specific paper-pencil tests to measure a small segment of GPK (Bouas, 1996; Sciotto et al., 2000) are related to one of the broader fields only. While cooperative learning as a teaching strategy is clearly related to the instructional process, the diagnosis of attention deficit disorder refers to student learning.

- video-based assessments to measure GPK can be broader thus covering all of the three content areas (König et al., 2014) or they are much more specific with regards to a certain teacher task such as classroom management (König & Lebens, 2012a) or a variety of aspects of the instructional process (Seidel et al., 2010).
- studies that intend to draw conclusions about teacher knowledge by rating the demonstration of teaching skills especially lesson planning (Pecheone & Chung, 2006; König, Buchholtz & Dohmen, under review) can be assigned to the field of instructional processes. Moreover, since the approach by König et al. (under review) also takes into account cognitive requirements and teaching strategies for diverse learners, student learning is also covered. The PACT assessment has a broad spectrum of contents, so in general all three fields seem to be covered.

Figure 12: Topics and areas covered in other studies reviewed in Part 2

		instructional process	student learning	assessment
Bouas, 1996	Knowledge of Cooperative Learning	x		
Sciutto, Terjesen, & Frank (2000)	KADDS		x	
König et al., 2014	TEDS-FU	x	x	x
König & Lebens, 2012a	CME	x		
Seidel, Blomberg & Stürmer, 2010	OBSERVE	x		
Zohar, 2004; see also Zohar, 2002; 2006 ***	Thinking in Science Classroom (TSC)			
Pecheone & Chung, 2006	Performance Assessment for California Teachers (PACT)	x	x	x
König, Buchholtz & Dohmen (under review)	PLANVOLL	x	x	

covered (x), difficult to assign (***)

Use for this review

188. To sum up, although defining content is a great challenge, three broader fields of GPK for teaching can be identified across all measurement instruments developed and applied in previous studies. Most likely, difficulties arise again on the level of operationalisation, i.e., when items are developed. Content of GPK can be defined at least from two sides: either from typical tasks teacher have to master (e.g., classroom management, teaching for diversity) or from academic disciplines typically underlying general pedagogy (e.g., educational psychology, didactics, sociology of education). To identify typical tasks teachers have to master, job analyses can be used (as done for the PRAXIS series, see Robustelli & Tannenbaum, 2010). To identify academic disciplines, reviews of existing information could be used (e.g., analysis of documents such as textbooks, curriculum frameworks, standards; as done in, e.g., the LEK-study, see Seifert & Schaper, 2012) or a Delphi-study could be carried out (as done in the BILWISS-study, see Kunina-Habenicht et al., 2013).

Cultural Sensitivity

Introduction to the Section

189. Educational philosophies and cultural habits vary across countries and may have an influence on schooling, pedagogy, and teacher education. For example, a comparative study of primary school education in five countries (England, France, India, Russia, and the US) worked out how culture and pedagogy are related in each country (Alexander, 2001). Proposing principles and a framework for a comparative pedagogy, Alexander (2009, p. 23) formulates pedagogy as “the act of teaching”, aside its discourses of values and ideas. He suggests a heuristic with which pedagogy using this definition can be framed for comparative analysis (Figure 13).

Figure 13: Pedagogy as ideas: theories, values, beliefs and justifications (Alexander, 2009, p. 17)

Classroom level: ideas which enable teaching:	
• Students	characteristics, development, motivation, needs, differences
• Learning	nature, facilitation, achievement and assessment
• Teaching	nature, scope, planning, execution and evaluation
• Curriculum	ways of knowing, doing, creating, investigating and making sense
System/policy level: ideas which formalise and legitimate teaching:	
• School	e.g. infrastructure, staffing, training
• Curriculum	e.g. aims, content
• Assessment	e.g. formal tests, qualifications, entry requirements
• Other policies	e.g. teacher recruitment and training, equity and inclusion
Cultural/societal level: ideas which locate teaching:	
• Culture	the collective ideas, values, customs and relationships which inform and shape a society’s view of itself, of the world and of education
• Self	what it is to be a person; how identity is acquired

190. Interestingly, the ideas that enable teaching on the classroom level (Figure 11) have their correspondence in the three broader fields that were identified in the previous section when contents of the existing GPK measurement instruments were compared. Obviously, when pedagogy is seen from a comparative perspective, student (student learning), teaching (instructional processes), and forms of assessment/evaluation are common denominators.

191. Moreover, it is crucial to consider pedagogy and thus an assessment of teachers’ GPK being dependent on its context. In Figure 11, it becomes clear, that besides the classroom context where a teacher applies his or her knowledge, the system/policy level and the cultural/societal level have great importance. Such levels and their specific elements have to be accounted for when conceptualising teachers’ GPK or interpreting comparative assessment data. For example, in TEDS-M, a broad study framework following the concept of opportunities to learn (Mc Donnel, 1995) was applied to collect data on the intended and the implemented curriculum of teacher education systems which then could be used to provide an appropriate context to interpret international assessment data of future teachers’ professional knowledge as an outcome of teacher education (cf. Tatto et al., 2008; 2012).

192. Although the shape of general pedagogy is strongly influenced by cultural perspectives on the means and ends of schooling and on the role of teachers (Hopmann & Riquarts, 1995), there is some commonality due to the nature of teaching. A literature review revealed that two tasks of teachers can be regarded as core tasks in almost all countries: instruction and classroom management (see König et al. 2011). 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 and nurturing students' social and moral development or knowledge about school management should be included in the area of general pedagogy.

193. Most likely, serious problems of cultural sensitivity do not occur when looking at declarative-conceptual GPK that belongs to basic knowledge a teacher has to know and that has long become part of a basic understanding of the teacher profession, e.g., since it has been outlined in course books of initial teacher education and that is part of typical courses in many teacher education programmes worldwide, for example because there is appropriate evidence showing this is knowledge that was generated in the disciplinary research discourse. Such knowledge can be denoted as theoretical-formal knowledge. In TEDS-M (König et al., 2011), for example, this was the case with the test item in Figure 3. None of the experts who reviewed the test material disagreed with the categorisation of “right” and “wrong”, although these experts came from three very different countries (the US, Taiwan, and Germany).

194. However, problems of cultural sensitivity can occur when GPK is tested in a way that goes beyond declarative-conceptual knowledge. The application of knowledge in a professional context can imply that the influence of normative expectations typical for the specific cultural context increases. For example, although agreement might exist a teacher should know the conceptual difference between a student's intrinsic vs. extrinsic motivation (as measured by the item in Figure 3), less agreement across cultural context might exist regarding the question which of these motivation is to be higher valued with regards to student learning in the classroom. Similar to this, for example in TEDS-M, developing internationally valid scoring schemes for open response items such as shown in Figure 4 and Figure 5 were extraordinary challenging. Several expert reviews in which experts from all participating countries participated were indispensable. The review processes showed some of the items had to be excluded from the comparative assessment for reasons of validity. However, in the end it was possible to find broad agreement among experts participating in this review process regarding those test items that remained in the test instrument, which showed that an international large-scale assessment of teachers' GPK using paper-and-pencil instruments is basically possible.

195. Problems of cultural sensitivity can also occur when the information provided in the item stem is not completely free of cultural contexts. In Figure 3, the statement “A student learns before a test in mathematics, because he/she...” can be met in a great variety of cultural contexts. By contrast, a video-based stimulus such as those used in the studies reviewed in the section “studies using a video-based stimulus for testing knowledge and/or skills” (König et al., 2014; König & Lebens, 2012a; Seidel et al., 2010) can hardly be transferred to another linguistic context, and even if subtitles are used watching a classroom situation taped in another country might put teachers being tested at a disadvantage, because they are not familiar with the specific habits (e.g., gestures of the teacher) or they are distracted by unfamiliar elements (e.g., school uniform) in the classroom. So it remains an open question how to carry out a valid video-based assessment in a cross-national study to measure teacher knowledge comparatively. As far as the author knows no study exists that has tried to measure teacher knowledge in a standardised way across countries using video-based item prompts.

Assessment Formats

Introduction to the Section

196. This section deals with the following questions: How can the relevant types or qualities of teacher knowledge be assessed?

197. As the previous Part 2 has shown, research studies investigating GPK of teachers use a variety of research methods. We have generally distinguished empirical studies that actually test teacher knowledge from phenomenological studies that aimed at exploring and describing knowledge of teachers. Whereas studies using standardised tests make use of quantitative methods allowing to state who has a “higher” or “lower” test score, phenomenological studies usually use qualitative methods with more degrees of freedom concerning statements about such an hierarchical assignment of teachers according to their data (see, for details, König, 2014). Thus the term “assessment” as it is used here will be related to the first group of studies. In the following the methods used by quantitative studies that test teacher knowledge are looked at in detail. They can be grouped into the following themes: paper-pencil, video-vignette, item types (cf., e.g., Schmeiser & Welch, 2006).

Paper-Pencil

198. Paper-pencil-tests are used in several studies reviewed (e.g., TEDS-M, COACTIV-R, PRAXIS). Such measurement instruments can contain different item types, usually distinguishing between multiple-choice response items and open-response items or short-answer constructed-response items (cf., e.g., Schmeiser & Welch, 2006).

199. Multiple-choice response items are easy to handle and probably a good way to assess declarative knowledge. Since they may be limited in scope with regards to the assessment of intellectual skills with which knowledge is applied (cf., e.g., Miller, 1990), open-response items or short-answer constructed-response items can be the better choice. Moreover, accounting for the challenge that the application of pedagogical knowledge happens in a professional situation (Putnam & Borko, 2000), test items are framed with context information. In TEDS-M, for example, this was provided with a short description of a typical classroom situation problem that had to be solved by generating concrete strategies for which (future) teachers need to retrieve information from long-term memory, but also knowledge linked with classroom situation experience. In the PRAXIS test, two case studies are provided to teacher candidates in order to account for the significance and complexity of situational aspects of teacher challenges.

Video-Vignette

200. Paper-pencil-tests have been scrutinised with regards to the validity of teacher knowledge measurement. It has been argued that they only assess “low-level or marginally relevant knowledge and skills” (Darling-Hammond & Baratz-Snowden, 2005, p. 61). Therefore some researchers use video clips of classroom situations as a stimulus, followed by test items related to the specific clip which teachers have to watch and respond to in order to bring about their professional knowledge.

201. Regarding the current state of research on measuring teacher knowledge using video-vignettes, it seems to be difficult to judge whether video-vignettes are generally superior to classical paper-pencil-tests. In the field of general pedagogical knowledge, there is, as far as the author knows, only one study that compares the two modes of measurement by conducting correlational analysis (König et al., 2014).

Item-Types

202. Usually, item types are classified into the various formats such as multiple-choice-response-items, open-response-items, constructed response items, complex constructed response items (for an overview see, e.g., Schmeiser & Welch, 2006). If studies use open-response items, they have to develop coding rubrics (see, e.g., König et al., 2011; 2014). Usually, an inter-rater reliability coefficient such as Cohen's Kappa for independently coding of items by two raters is used to examine the quality of the coding schemes (Jonsson & Svingby, 2007).

203. Besides coding-schemes, it is essential to define which code stands for a correct, a partially correct or a false answer. While regarding the measurement of declarative knowledge this is not such a complex task, test items measuring knowledge that is of situated nature or procedural knowledge sometimes require a very time-consuming development process of scoring schemes. Taking the test item in Figure 4 as an example, it was a long procedure through several expert reviews to agree on the definition of four criteria to be used in order to score future teachers' responses (as exemplarily presented in Figure 6).

PART 4. CONCLUSION AND RECOMMENDATIONS

204. This part contains conclusions drawn from what has been described and analysed in the previous parts. We aim to conclude with recommendations resulting from these considerations.

Content Areas to be Covered and Types of Knowledge to be Assessed

Content Areas

205. As outlined in Part 3, to identify the content of teacher GPK is a challenging task. Together with what has been discussed and proposed in the literature one can use typical tasks teachers have to master (“task-based approach”) or one can use academic disciplines underlying general pedagogy (“disciplinary approach”) as a starting point and then derive a definition of GPK to be tested. Very interestingly, the approaches applied by the studies reviewed in Part 2 vary to a certain degree, but they are not so different from each other. As could be seen in the discussion of conceptual issues in Part 3, there is overlap in the three broader fields, at least when looking on teaching as the core task of teachers:

- *instructional process* (including teaching methods, didactics, structuring a lesson, and classroom management);
- *student learning* (including their cognitive, motivational, emotional individual dispositions; their learning processes and development; their learning as a group taking therefore into account student heterogeneity and adaptive teaching strategies); and
- *assessment* (including diagnosing principles irrespective of the subject, evaluation procedures).

So it seems there is some agreement regarding what content is relevant for the definition of GPK.

206. However, one principle problem has to be mentioned here. Both approaches applied (task-based, disciplinary) most likely reflect what general pedagogy currently *is*. Less attention may be given to what general pedagogy *should be*. So assessment instruments reviewed here predominantly apply a *descriptive* way to assess GPK, i.e., they either use existing (observable) teacher tasks or they use existing (institutionalised) academic disciplines to provide reasons and arguments for why they have included or excluded certain content of general pedagogy. This is, in any case, a legitimate and solid way to define test content (cf. Schmeiser & Welch, 2006). Nevertheless, the crucial question arises whether innovative aspects to define GPK tend to be neglected. Especially with regards to a new understanding of teachers’ pedagogical knowledge influenced by the research of mind, brain, and education (OECD working paper by Ansari, König, Leask, & Tokahoma), there is the important question of how to expand hitherto definitions of teachers’ GPK.

Types of Knowledge

207. Teacher Knowledge is heterogeneous, and so various types of teacher knowledge have to be distinguished. The studies reviewed provide a good basis to becoming more precise on this issue. A central issue is that teacher knowledge is multidimensional. It does not only consist of declarative knowledge directly derived from the formal structure of an academic discipline. Instead, various forms and qualities are typical for a profound description of the types of teacher knowledge.

208. Some of the studies reviewed are committed to assess performative aspects of teachers' GPK. Whether they use complex open-response test items in paper-pencil tests, video-cued test items, or an analysis of artefacts such as authentic lesson plans written for a real learning group, the common denominator of such efforts is there is agreement on the issue that higher order knowledge of teachers is of a situated nature (Putnam & Borko, 2000) and dependent on the specific situation in the context of the learning group or the classroom. Even if such situated or procedural knowledge itself is dependent on factual knowledge, it is very difficult to assess differences between novice and expert teachers' GPK when only focusing on declarative knowledge of GPK.

209. To conclude, a profound assessment of teachers' GPK has to include test items measuring declarative knowledge, but in addition to that, it has to go beyond that type of knowledge (Shavelson, 2010). Performative aspects such as the application of that declarative knowledge in a variation of specific situations that are typical for the teaching profession have to be accounted for. From the research discourse one may argue that only then an assessment of teachers' GPK is a valid testing of teachers' competence. From a more practical perspective, one may assume that only then will be an assessment of teachers' GPK evoke broad acceptance among in-service teachers, educational policy makers and other representatives of the everyday, practical in-school context.

Methodological Approach

Introduction to the Section

210. Tests are usually developed for different purposes and thus follow different methodological approaches and standards. According to Tittle (2006), for example, there are three different purposes of tests related to the target group of teachers: measurement for licensure, measurement for teacher development, and measurement for research.

211. In this section we lay a special focus on methodological approaches that come with considerations on the challenge of designing an international instrument to assess teachers' general pedagogical knowledge. However, it is not the purpose of this paper to provide general information on international large-scale assessments or educational measurement (see for further reading, e.g., National Research Council, 2002; Brennan, 2006; Ross & Genevois, 2006; Rutkowski et al., 2013).

Test Design and Construct Validity

212. It is important to examine construct validity of measures that are used by studies investigating teachers' GPK. This should include classroom observations of teacher performance and student achievement. Although progress has been made regarding the development of measurement instruments assessing GPK as the studies reviewed show there are very few studies that have already applied such measurement instruments to predict such outcomes of teacher knowledge. Critically seen, there is a tremendous research deficit in this area, so future research studies should be committed to validate GPK tests against such outcomes in order to prove the quality of their assessment.

Length

213. Large-scale assessments usually make use of an item rotation design. In the tests reported here, this was applied in several studies (such as TEDS-M, LEK, EMW, DIDAKTUM, TEDS-FU, BILWISS). Using a balanced incomplete block design (Adams & Wu, 2002; von Davier, Carstensen, & von Davier, 2006) with a certain number of test booklets means each person has to respond to only a selection of all test items. The advantage is that the construct can be measured by a broad range of test items, while the single person who is being tested is less burdened with the amount of test items.

214. Item response theory (IRT) scaling methods can be used to estimate scores across the different booklets. With the methods implemented in a software package like Conquest (Wu, Adams, & Wilson, 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. So, very often, items have to be fairly equally distributed across content dimensions of a test. That approach, however, needs a number of pilot studies in order to assure evidence-based decisions about test design and test booklet assembly.

Proficiency Scaling

215. Usually, test scores provide information about the degree to which an individual's performance matches with the scale of the assessment in relation to other individuals. Some tests report overall scores and subscores of a test design matrix. In addition to this, proficiency scaling is a relevant strategy to generate criterial information about test scores.

216. To sum up, test items are grouped according to certain characteristics, such as item difficulty, quality of cognitive demand, or conceptual commonalities. Then using this grouping, cut-off values on the test score continuum are defined to allow the application of these cut-off values to the person's ability estimate. As a result, groups of items can be assigned to groups of persons with a test performance that can be described by the characteristics (e.g., item difficulty) of the grouped test items. E.g., the measurement of reading literacy in PISA contains five or six (depending on the cycle) proficiency levels.

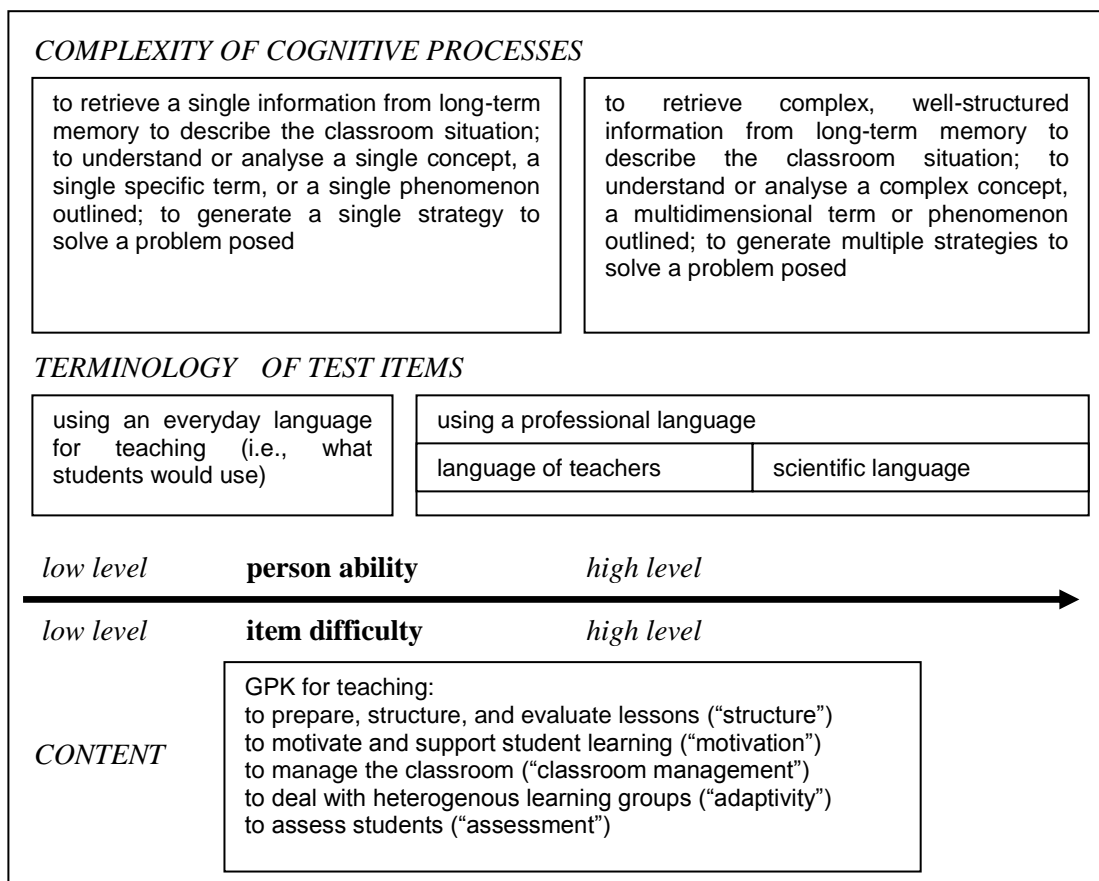
217. Regarding the measurement of teachers' GPK, proficiency scaling seems to be a relevant issue, e.g., when informing readers with criteria-related description of test scores. Looking at the studies reviewed, criteria-related descriptions are provided by, e.g., the Praxis series assessment instrument or the PACT. However, they are not proficiency scaling in the closer sense, but they are "score profiles". For example, in the PACT assessment, the 4-point Likert scale continuum is defined as follows (Pecheone & Chung, 2006, p. 25): "A score of 1 on an individual GQ [guiding question] means the candidate has not met the standard on that item. A score of 2 means the candidate has met the standard at the minimum level. Scores of 3 and 4 represent advanced and superior levels of performance, respectively."

218. In a study that used the TEDS-M instrument to measure GPK, a strategy how proficiency levels of first phase future teachers' pedagogical knowledge was modelled using psychometric analysis (König, 2009). The TEDS-M test instrument measuring pedagogical knowledge was applied to a sample of 197 students at the University of Cologne. The *terminology of test items* (everyday language, professional language, educational-scientific language) as well as *the complexity of cognitive processes* (regarding, e.g., the complexity of strategies how to deal with a typical problem in the classroom) were used as two criteria to describe test item difficulty thus defining groups of items and were incorporated into a corresponding measurement model (Figure 14). The assignment of items to groups was done first on a theoretical level, then verified by an expert review (see for details, König, 2009). Item parameter estimation distributions differentiated by the criteria used to describe test difficulty were nearly as expected and enabled the definition of proficiency levels. Frequency distributions of future teachers differentiated by teacher education programme and period of training pointed out the content validity of the modelled proficiency levels.

219. Although future research is needed to investigate proficiency scaling of teachers' GPK in depth, the study by König (2009) shows that basically it is possible to carry out proficiency scaling of teachers' GPK. Therefore, as early as when the development of a new assessment instrument is launched, one should clarify what will be defined as low, medium, and high level ability and how these various levels will be measured by what kind of test items that would then in turn provide concrete criteria to give a description of the competence being measured at each level. This is the core of competence measurement modelling

(see, e.g., the PISA reading literacy model as one of the best examples hitherto created) and should be applied to the measurement of teacher knowledge in future studies.

Figure 14: Competence model proposed by König (2009, p. 248) using the TEDS-M instrument and integrating GPK content, cognitive demands, and explicit item difficulties (terminology of test items, complexity of cognitive processes)



REFERENCES

- Adams, R., & Wu, M. (Eds.). (2002). *PISA 2000 technical report*. Paris, France: OECD. Retrieved from <http://www.oecd.org/dataoecd/53/19/33688233.pdf>
- Alexander, R. (2001). *Culture and Pedagogy: international comparisons in primary education*. New York/Oxford: Wiley/Blackwell Sci.
- Alexander, R. (2009). Pedagogy, culture and the power of comparison. In H. Daniels, H. Lauder, & J. Porter (eds.), *Educational theories, culture and learning: A critical perspective* (pp. 10-26). London: Routledge.
- Allen, R. M., & Casbergue, R. M. (1997). Evolution of novice through expert teachers' recall: Implications for effective reflection on practice. *Teaching and Teacher Education*, 13, 741–755.
- Anderson, J. R. (1982). Acquisition of cognitive skills. *Psychological Review*, 89, 369–406.
- 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, NY: Longman.
- Atjonen, P., Korkeakoski, E., & Mehtäläinen, J. (2011). Key pedagogical principles and their major obstacles as perceived by comprehensive school teachers. *Teachers and Teaching: theory and practice*, 17(3), 273-288.
- Baer, M. et al. (2011). *Lehrerbildung und Praxiserfahrung im ersten Berufsjahr und ihre Wirkung auf die Unterrichtskompetenzen von Studierenden und jungen Lehrpersonen im Berufseinstieg* [Teacher training and practical experiences in the first year of work and their effects on the teaching competencies of students and teachers at the start of their career]. *Zeitschrift für Erziehungswissenschaft*.
- 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.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9, 469-520.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A. et al. (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. *International Journal of Educational Research*, 35(5), 463-482.
- Berliner, D.C. (2004). Describing the behavior and documenting the accomplishments of expert teachers. *Bulletin of Science, Technology and Society*, 24(3), 200-212.
- Blömeke, S., & Delaney, S. (2012). Assessment of teacher knowledge across countries: A review of the state of research. *ZDM*, 44(3), 223-247.

- Blömeke, S., Kaiser, G. & König, J. (2009). Längsschnittliche Entwicklung der Kompetenzen von Junglehrkräften: Follow-Up zur internationalen Vergleichsstudie TEDS-M (TEDS-FU). Antrag auf Gewährung einer Sachbeihilfe an die Deutsche Forschungsgemeinschaft.
- Blömeke, S. & König, J. (2011). Zum Zusammenhang von Ausbildungsformen, -inhalten und -methoden mit dem erworbenen pädagogischen Professionswissen von Grundschullehrkräften. *Zeitschrift für Grundschulforschung*, 4(1), 33-46.
- Blömeke, S., & König, J. (2010). Pädagogisches Wissen angehender Mathematiklehrkräfte im internationalen Vergleich [Pedagogic knowledge of prospective mathematics teachers in an international comparison]. In S. Blömeke, G. Kaiser, & R. Lehmann (Eds.), *TEDS-M 2008—Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte im internationalen Vergleich [TEDS-M 2008—Professional competence and learning opportunities of prospective mathematics teachers in an international comparison]* (pp. 270-283). Münster, Germany: Waxmann.
- Blömeke, S., Paine, L., Houang, R. T., Hsieh, F. J., Schmidt, W. H., Tatto, M. T., ... & Schwille, J. (2008). Future teachers' competence to plan a lesson: first results of a six-country study on the efficiency of teacher education. *ZDM*, 40(5), 749-762.
- Borko, H. & Putnum, R.T. (1996). Learning to Teach. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 673-707). New York: Simon & Schuster.
- Bouas, M. J. (1996). Are we giving cooperative learning enough attention in preservice teacher education? *Teacher Education Quarterly*, 23, 45-58.
- Brennan, R.L. (ed.). *Educational Measurement* (4th ed.). ACE.
- Bromme, R. (1992). *Der Lehrer als Experte: zur Psychologie des professionellen Wissens [The teacher as expert: The psychology of professional knowledge]*. Bern, Switzerland: Huber.
- Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In Weinert, F.E. (Eds.), *Enzyklopädie der Psychologie: Psychologie des Unterrichts und der Schule*. Bd. 3 (pp. 177-212). Göttingen: Hogrefe.
- Bromme, R. (2001). Teacher Expertise. *International Encyclopedia of the Social & Behavioral Sciences* (pp. 15459-15465). Amsterdam: Smelser and Baltès.
- Brown, M. B., Bolen, L. M., Lassiter, C. L., & Burke, M. M. (2006). Competencies of Traditionally Prepared and Lateral Entry Teachers: Implications for School Administrators. *AASA Journal of Scholarship & Practice*, 3(1), 18-25.
- Carter, K., Sabers, D.S., Cushing, K.S., Pinnegar, P. & Berliner, D.C. (1987). Processing and using information about students: A study of expert, novice and postulant teachers. *Teaching and Teacher Education*, 3, 147-157.
- Choy, D., Chong, S., Wong, A. F., & Wong, I. Y. F. (2011). Beginning teachers' perceptions of their levels of pedagogical knowledge and skills: did they change since their graduation from initial teacher preparation?. *Asia Pacific Education Review*, 12(1), 79-87.

- Choy, D., Lim, K. M., Chong, S., & Wong, A. F. (2012). A confirmatory factor analytic approach on Perceptions of Knowledge and Skills in Teaching (PKST) 1. *Psychological Reports*, 110(2), 589-597.
- Clarridge, P. B., & Berliner, D. C. (1991). Perceptions of student behavior as a function of expertise. *Journal of Classroom Interaction*, 26, 1-8.
- Darling-Hammond, L. (2000). Teacher quality and student achievement. A review of state policy evidence. *Education Policy Analysis Archives*, 8.
- De Jong, T., & Ferguson-Hessler, M. G. (1996). Types and qualities of knowledge. *Educational psychologist*, 31(2), 105-113.
- Douville, P., Pugalee, D. K., & Wallace, J. D. (2003). Examining instructional practices of elementary science teachers for mathematics and literacy integration. *School Science and Mathematics*, 103(8), 388-396.
- Doyle, W. (1985). Recent Research on Classroom Management: Implications for Teacher Preparation. *Journal of Teacher Education*, 36, 31-35.
- Doyle, W. (2006). Ecological Approaches to Classroom Management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 97-125). Mahwah, NJ: Erlbaum.
- ETS (2010). *Praxis™ Technical Manual*. Revised December 2010. Princeton, New Jersey: Educational Testing Service.
- ETS (2011). *Multi-State Standard Setting Technical Report. Praxis Principles of Learning and Teaching: Grades 5-9 (0623)*. Princeton, New Jersey: Educational Testing Service.
- ETS (2013). *Study Guide for Principles of Learning and Teaching: Grades 5-9 and 7-12*. Princeton, New Jersey: Educational Testing Service.
- Garrahy, D. A., Cothran, D. J., & Kulinna, P. H. (2005). Voices from the trenches: An exploration of teachers' management knowledge. *Journal of Educational Research*, 99, 56-63. doi:10.3200/JOER.99.1.56-63
- Garza, R. & Ovando, M. (2012). Preservice teachers' connections of pedagogical knowledge to mentoring at-risk adolescents: Benefits and challenges. *Partnership in Learning*, 20(3), 343-360.
- Gatbonton, E. (1999). Investigating experienced ESL teachers' pedagogical knowledge. *The Modern Language Journal*, 83(1), 35-50.
- Gonzalez, L. E., & Carter, K. (1996). Correspondence in cooperating teachers' and student teachers' interpretations of classroom events. *Teaching and Teacher Education*, 12, 39-47.
- Grossman, P. L. (1990). *The making of a teacher*. New York, NY: Teachers College Press.
- Grossman, P. L. (1992). Why models matter: An alternate view on professional growth in teaching. *Review of Educational Research*, 62(2), 171-179

- Hativa, N. (2000). Becoming a better teacher: A case of changing the pedagogical knowledge and beliefs of law professors. *Instructional Science*, 28(5), 491-523.
- Hattie, J. (2009) *Visible Learning: a synthesis of over 800 meta-analyses relating to achievement*. London, Routledge.
- Hattie, J. A. C. (2012). *Visible Learning for Teachers: Maximizing impact on learning*. London, New York. Routledge.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42(2), 371-406.
- Hogan, T., Rabinowitz, M., & Craven III, J. A. (2003). Representation in teaching: Inferences from research of expert and novice teachers. *Educational Psychologist*, 38(4), 235-247.
- Housner, L.D. & Griffey, D.C. (1985). Teacher cognition: Differences in planning and interactive decision-making between experienced and in-experienced teachers. *Research Quarterly for Exercise and Sport*, 56, 45-53.
- 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.
- Kaiser, G., Busse, A., Benthien, J., König, J. & Blömeke, S. (under review). About the complexities of video-based assessments: Theoretical and methodological approaches to overcoming shortcomings of teacher research. *International Journal of Science and Mathematics Education*.
- Kansanen, P. (2002). Didactics and its relation to educational psychology: Problems in translating a key concept across research communities. *International Review of Education*, 48(6), 427-441.
- Kansanen, P., Tirri, K., Meri, M., Krokfors, L., Husu, J., & Jyrhämä, R. (2000). *Teacher's pedagogical thinking*. New York: Peter Lang.
- Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S. & Baumert, J. (2013). Teachers' Content Knowledge and Pedagogical Content Knowledge. The Role of Structural Differences in Teacher Education. *Journal of Teacher Education*, 64(1), 90-106.
- Klette, K., & Bergem, O. Kr. (2012). *Considerations for a TALIS Video Study*. University of Oslo, Norway. Paris: OECD Working Papers.
- Klieme, E. (2012). Internationales large scale assessment in der Lehrerbildung: Anmerkungen zu einem neuen Paradigma der vergleichenden Bildungsforschung. *Zeitschrift für Pädagogik*, 58, 492-499.
- KMK (2004). *Standards für die Lehrerbildung: Bildungswissenschaften*. Beschluss der Kultusministerkonferenz. Bonn/Germany.

- Kompf, M., & Denicolo, P. (2003). *Teacher thinking twenty years on: Revisiting persisting problems and advances in education*. Lontoo: Taylor & Francis.
- König, J. (2009). Zur Bildung von Kompetenzniveaus im Pädagogischen Wissen von Lehramtsstudierenden: Terminologie und Komplexität kognitiver Bearbeitungsprozesse als Anforderungsmerkmale von Testaufgaben? [On the modelling of proficiency levels of future teachers' pedagogical knowledge: Can test item terminology and complexity of cognitive processes be used to describe item difficulty?] *Lehrerbildung auf dem Prüfstand*, 2(2), 244- 262.
- König, J. (2013). First comes the theory, then the practice? On the acquisition of general pedagogical knowledge during initial teacher education. *International Journal of Science and Mathematics Education*, 11(4), 999-1028. [<http://dx.doi.org/10.1007/s10763-013-9420-1>]
- König, J. (2014). Kompetenzen in der Lehrerbildung aus fächerübergreifender Perspektive der Bildungswissenschaften [Competences in Teacher Education from a Generic Perspective of Educational Science]. In A. Bresges, B. Dilger, T.Hennemann, J. König, H. Lindner, A. Rohde, & D. Schmeinck (Eds.), *Kompetenzen diskursiv. Terminologische, exemplarische und strukturelle Klärungen in der LehrerInnenbildung [Competence Discourse. Terminological, Exemplifying, and Structural Clarifications in Teacher Education]* (pp. 15-46). Münster: Waxmann.
- König, J. & Blömeke, S. (2009a). Pädagogisches Wissen von angehenden Lehrkräften: Erfassung und Struktur von Ergebnissen der fachübergreifenden Lehrerausbildung [Pedagogic knowledge of future teachers. Capturing and structuring the results of cross-subject teacher education]. *Zeitschrift für Erziehungswissenschaft*, 12(3), 499-527.
- König, J. & Blömeke, S. (2009b). Pädagogisches Wissen von österreichischen Lehramtsstudierenden [Pedagogical Knowledge of Austrian Student Teachers]. *Erziehung & Unterricht*, 159(1/2), 175-186.
- König, J. & Blömeke, S. (2012). Future Teachers' General Pedagogical Knowledge from a Comparative Perspective. Does School Experience Matter? *ZDM - The International Journal on Mathematics Education*, 44(3), 341-354. [<http://rd.springer.com/article/10.1007/s11858-012-0394-1>]
- König, J. & Blömeke, S. (2013). Preparing Teachers of Mathematics in Germany. In J. Schulle, L. Ingvarson & R. Holdgreve-Resendez (eds.), *TEDS-M Encyclopaedia. A Guide to Teacher Education Context, Structure and Quality Assurance in 17 Countries. Findings from the IEA Teacher Education and Development Study in Mathematics (TEDS-M)* (pp. 100-115). Amsterdam: IEA.
- König, J., Blömeke, S., & Kaiser, G. (under review). Early career mathematics teachers' general pedagogical knowledge and skills: Do teacher education, teaching experience, and working conditions make a difference? *International Journal of Science and Mathematics Education*.
- König, J., Blömeke, S., Klein, P., Suhl, U., Busse, A., & Kaiser, G. (2014). Is teachers' general pedagogical knowledge a premise for noticing and interpreting classroom situations? A video-based assessment approach. *Teaching and Teacher Education*, 38, 76-88
- König, J., Blömeke, S., Paine, L., Schmidt, B. & 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, 188-201.
- König, J., Blömeke, S., Paine, L., Schmidt, W.H. & Hsieh, F.-J. (2014). Teacher education effectiveness: Quality and equity of future primary and future lower secondary teachers' general pedagogical knowledge. In S. Blömeke et al. (eds.), *International Perspectives on Teacher Knowledge, Beliefs,*

and Opportunities to Learn. *Advances in Mathematics Education* (pp. 187-206). Dordrecht: Springer Science + Business Media.

König, J., Buchholtz, C. & Dohmen, D. (under review). Analyse von schriftlichen Unterrichtsplanungen: Empirische Befunde zur didaktischen Adaptivität als Aspekt der Planungskompetenz angehender Lehrkräfte [Lesson plan analysis: Empirical findings on pedagogical adaptivity as an aspect of future teachers' competence to plan a lesson]. *Zeitschrift für Erziehungswissenschaft*.

König, J., Kaiser, G. & Felbrich, A. (2012). Spiegelt sich pädagogisches Wissen in den Kompetenzselbsteinschätzungen angehender Lehrkräfte? Zum Zusammenhang von Wissen und Überzeugungen am Ende der Lehrerausbildung [Is Pedagogical Knowledge Reflected in the Competence-Related Self-Assessments of Future Teachers? On the inter-relation between knowledge and beliefs after completing teacher training]. *Zeitschrift für Pädagogik*, 58(4), 476-491.

König, J., & Lebens, M. (2012a). Classroom Management Expertise (CME) von Lehrkräften messen: Überlegungen zur Testung mithilfe von Videovignetten und erste empirische Befunde. [Measuring teachers' Classroom Management Expertise (CME): On the testing via video-vignettes and first empirical findings.] *Lehrerbildung auf dem Prüfstand*, 5(1), 3-29.

König, J., & Lebens, M. (2012b). Fragebogen zur Erfassung von Classroom Management Expertise (CME) über vier Video-Vignetten [Test Items on the Measurement of Classroom Expertise of Teachers (CME) via Four Video-Vignettes]. University of Cologne: Institute for General Didactics and School Research.

König, J., & Pflanzl, B. (under review). On the relationship between teachers' general pedagogical knowledge and instructional quality.

König, J., & Rothland, M. (2012). Motivations for Choosing Teaching as a Career: Effects on General Pedagogical Knowledge during Initial Teacher Education. *Asia-Pacific Journal of Teacher Education*, 40(3), 291-317.

König, J., & Rothland, M. (under review). Wissen über Klassenführung als Ressource der Burnout-Prävention? Zum Nutzen von pädagogisch-psychologischem Wissen im Lehrerberuf [Knowledge of Classroom Management as a Resource for Preventing Burnout? On the Use of Pedagogical-Psychological Knowledge in the Teaching Profession]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*.

König, J., Rothland, M., Darge, K., Lünemann, M. & Tachtsoglou, S. (2013). Erfassung und Struktur berufswahlrelevanter Faktoren für die Lehrerausbildung und den Lehrerberuf in Deutschland, Österreich und der Schweiz [Detection and structure of career choice factors for teacher training and the teaching profession in Germany, Austria and Switzerland]. *Zeitschrift für Erziehungswissenschaft*, 16 (3), 553-577. <http://dx.doi.org/10.1007/s11618-013-0373-5>

König, J., & Seifert, A. (Hrsg.) (2012). Lehramtsstudierende erwerben pädagogisches Professionswissen. Ergebnisse der Längsschnittstudie LEK zur Wirksamkeit der erziehungswissenschaftlichen Lehrerausbildung [Student Teachers Acquire Professional Pedagogical Knowledge. Findings from Longitudinal Study LEK on the Effectiveness of General Pedagogy in Teacher Education]. Münster: Waxmann.

König, J., Tachtsoglou, S., Darge, K. & Lünemann, M. (2014). Zur Nutzung von Praxis: Modellierung und Validierung lernprozessbezogener Tätigkeiten von angehenden Lehrkräften im Rahmen ihrer schulpraktischen Ausbildung [On the use of practice: modeling and validating future teachers']

learning process related activities during in-school opportunities to learn]. *Zeitschrift für Bildungsforschung*, 4(1), 3-22. <http://dx.doi.org/10.1007/s35834-013-0084-2>

- 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.
- Kunina-Habenicht, O., Lohse-Bossenz, H., Kunter, M., Dicke, T., Förster, D., Gößling, J., Schulze-Stocker, F., Schmeck, A., Baumert, J., Leutner, D. & Terhart, E. (2012). Welche bildungswissenschaftlichen Inhalte sind wichtig in der Lehrerbildung? [Which educational topics are important for teacher training? Results of a Delphi study]. *Zeitschrift für Erziehungswissenschaft*, 15(4), 649-682.
- Kunina-Habenicht, O., Schulze-Stocker, F., Kunter, M., Baumert, J., Leutner, D., Förster, D., Lohse-Bossenz, H. & Terhart, E. (2013). Die Bedeutung der Lerngelegenheiten im Lehramtsstudium und deren individuelle Nutzung für den Aufbau des bildungswissenschaftlichen Wissens [The significance of learning opportunities in teacher training courses and their individual use for the development of educational-scientific knowledge]. *Zeitschrift für Pädagogik*, 59(1), 1-23.
- Larcher, S., & Oelkers, J. (2004). Deutsche Lehrerbildung im internationalen Vergleich [German teacher education: An international comparison]. In S. Blömeke, P. Reinhold, G. Tulodziecki, & J. Wildt (Eds.), *Handbuch Lehrerbildung* [Handbook of teacher training] (pp. 128-150). Bad Heilbrunn, Germany: Klinkhardt.
- Metzler, M. W., Tjeerdsma, B. J. & Mozen, D.M. (2000). Assessing pedagogical knowledge. *Journal of Teaching in Physical Education*, 19 487-507.
- Miller, G. E. (1990). The Assessment of Clinical Skills/Competence/Performance. *Academic Medicine. Journal of the Association of American Medical Colleges*, 65, 63-67.
- National Research Council (2002). *Methodological Advances in Cross-National Surveys of Educational Achievement*. Washington, DC: National Academy Press.
- O'Connor, E. A., & Fish, M. C. (1998, April). Differences in the classroom systems of expert and novice teachers. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- OECD (2005). *Teachers matter. Attracting, developing and retaining effective teachers*. Paris: OECD.
- O'Hara, S., & Pritchard, R. H. (2008). Meeting the Challenge of Diversity: Professional Development for Teacher Educators. *Teacher Education Quarterly*, 35(1), 43-61.
- Oser, F., & Oelkers, J. (Eds.). (2001). *Die Wirksamkeit der Lehrerbildungssysteme* [The effectiveness of teacher education systems]. Chur, Switzerland: Ruedger.
- Pecheone, R. L., & Chung, R. R. (2006). Evidence in teacher education: The Performance Assessment for California Teachers (PACT). *Journal of Teacher Education*, 57(1), 22-36.
- Pecheone, R. L., & Chung, R. R. (2007). *PACT technical report*. Stanford, CA: PACT Consortium.
- PACT Consortium (2004). *Beginning Teacher Quality: What Matters for Student Learning? A Research Proposal from Stanford University*.

- Pflanzl, B., Thomas, A., & Matischek-Jauk, M. (2013). Pädagogisches Wissen und pädagogische Handlungskompetenz [Pedagogical knowledge and pedagogical action competence]. *Erziehung und Unterricht*, 1/2, 1-8.
- 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.
- Robustelli, S. L. & Tannbenbaum, R. J. (2010). *Validity Evidence to Support the Development of a Licensure Assessment for Entry-Level Teachers: A Job-Analytic Approach*. Princeton, New Jersey: Educational Testing Service.
- Ross, K.N., & Genevois, I.J. (eds.) (2006). *Cross-national studies of the quality of education. Planning their design and managing their impact*. Paris: International Institute for Educational Planning.
- Rutkowski, L., von Davier, M. & Rutkowski, D. (2013). *Handbook of International Large-Scale Assessment: Background, Technical Issues, and Methods of Data Analysis*. CRC Press.
- Sabers, D. S., Cushing, K. S., & Berliner, D. C. (1991). Differences among teachers in a task characterized by simultaneity, multidimensionality, and immediacy. *American Educational Research Journal*, 28(1), 63–88.
- Sánchez, E., Rosaes, J., & Cañedo, I. (1999). Understanding and communication in expositive discourse: An analysis of the strategies used by expert and pre-service teachers. *Teaching and Teacher Education*, 15, 37–58.
- Sandholtz, J. H., & Shea, L. M. (2012). Predicting Performance A Comparison of University Supervisors' Predictions and Teacher Candidates' Scores on a Teaching Performance Assessment. *Journal of Teacher Education*, 63(1), 39-50.
- Schleicher, A. (2011). *Building a High-Quality Teaching Profession. Lessons from around the world*. Paris: OECD.
- Schmeiser, C.B., & Welch, C.J. (2006). Test Development. In R.L. Brennan (ed.), *Educational Measurement* (4th ed., pp. 307-353). ACE.
- Schmidt, W. H., Tatto, M. T., Bankov, K., Blömeke, S., Cedillo, T., Cogan, L., Han, S.-I., Houang, R., Hsieh, F.-J., Paine, L., Santillan, M. N. & Schwillie, J. (2007). *The preparation gap: Teacher education for middle school mathematics in six countries – Mathematics teaching in the 21st century (MT21 Report)*. http://usteds.msu.edu/related_research.asp
- Schmidt, W. H., Cogan, L., & Houang, R. (2011). The role of opportunity to learn in teacher preparation: An international context. *Journal of Teacher Education*, 62(2), 138-153.
- Schulte, K., Bögeholz, S., & Watermann, R. (2008). Selbstwirksamkeitserwartungen und pädagogisches Professionswissen im Verlauf des Lehramtsstudiums [Teacher self-efficacy and general pedagogical knowledge during the teacher education]. *Zeitschrift für Erziehungswissenschaft*, 11, 268–287.
- Sciutto, M. J., Terjesen, M. D., & Frank, A. S. B. (2000). Teachers' knowledge and misperceptions of attention-deficit/hyperactivity disorder. *Psychology in the Schools*, 37, 115–122.
- Seidel, T., Blomberg, G., & Stürmer, K. (2010). „Observer“ - validation of a video-based instrument to assess the professional view on classrooms. *Zeitschrift für Pädagogik*, 56. Beiheft, 296-30.

- Seidel, T., Prenzel, M., Schwindt, K., Stürmer, K., Blomberg, G., & Kobarg, M. (2009). LUV and Observe: Two Projects Using Video to Diagnose Teacher Competence. *The power of video studies in investigating teaching and learning in the classroom*, 243.
- Seifert, A. & König, J. (2012). Pädagogisches Unterrichtswissen - bildungswissenschaftliches Wissen: Validierung zweier Konstrukte. [Pedagogical knowledge for teaching – educational scientific knowledge: validating two constructs] In J. König & Seifert, A. (Hrsg.), *Lehramtsstudierende erwerben pädagogisches Professionswissen. Ergebnisse der Längsschnittstudie LEK zur Wirksamkeit der erziehungswissenschaftlichen Lehrerbildung* (S. 215-233). Münster: Waxmann.
- Seifert, A. & Schaper, N. (2012). Die Entwicklung von bildungswissenschaftlichem Wissen. In J. König & Seifert, A. (Hrsg.), *Lehramtsstudierende erwerben pädagogisches Professionswissen. Ergebnisse der Längsschnittstudie LEK zur Wirksamkeit der erziehungswissenschaftlichen Lehrerbildung* (S. 183-214). Münster: Waxmann.
- 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.
- Swanson, H. L., O'Connor, J. E., & Cooney, J. B. (1990). An information processing analysis of expert and novice teachers' problem solving. *American Educational Research Journal*, 27, 533-556.
- Tan, A. L., Tan, S. C., & Wettasinghe, M. (2011). Learning to be a science teacher: Reflections and lessons from video-based instruction. *Australasian Journal of Educational Technology*, 27(3), 446-462.
- 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, MI: Michigan State University.
- Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Rowley, G., Peck, R., Bankov, K., Rodriguez, M., & Reckase, M. (2012). Policy, practice, and readiness to teach primary and secondary mathematics in 17 countries. Findings from the IEA teacher education and development study in mathematics (TEDS-M). http://teds.educ.msu.edu/wp-content/uploads/IEA_TEDS-M-International-Report1.pdf
- Terhart, E., Schulze, F., Kunina-Habenicht, O., Dicke, T., Förster, D., Lohse-Bossenz, H., & Leutner, D. (2012). *Bildungswissenschaftliches Wissen und der Erwerb professioneller Kompetenz in der Lehramtsausbildung—Eine Kurzdarstellung des BilWiss-Projekts*. *Lehrerbildung auf dem Prüfstand*, 5(1), 96-106.
- Tittle, C.K. (2006). Assessment of teacher learning and development. In P. A. Alexander & P. H. Winne (eds.), *Handbook of educational psychology* (2nd ed., pp. 953-984). Mahwah, NJ: 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, Germany: Hogrefe.
- Voss, T., Kunter, M. & Baumert, J. (2011). Assessing Teacher Candidates' General Pedagogical/Psychological Knowledge: Test Construction and Validation. *Journal of Educational Psychology*, 103(4), 952-969.

- Wasson, B. (1998). Facilitating dynamic pedagogical decision making: PEPE and GTE. *Instructional Science*, 26(3-4), 299-316.
- Watt, H.M.G., Richardson, P.W., Klusmann, U., Kunter, M., Beyer, B., Trautwein, U., & Baumert, J. (2012). Motivations for choosing teaching as a career: An international comparison using the FIT-Choice scale. *Teaching and Teacher Education*, 28, 791–805.
- Wilson, S. M., Shulman, L. S., & Richert, A. E. (1987). “150 different ways” of knowing: Representations of knowledge in teaching. In J. Calderhead (Ed.), *Exploring teachers’ thinking* (pp. 104-124). London, UK: Cassell.
- Wilson, S. M., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. *Review of Research in Education*, 24, 173-210.
- Wu, M. L., Adams, R. J., & Wilson, M. R. (1997). *ConQuest: Multi-aspect test software* [computer program]. Camberwell, Australia: Australian Council for Educational Research.
- Zohar, A. (2004). Elements of teachers' pedagogical knowledge regarding instruction of higher order thinking. *Journal of Science Teacher Education*, 15(4), 293-312.
- Zohar, A. (2006). The nature and development of teachers' metastrategic knowledge in the context of teaching higher order thinking. *The journal of the learning sciences*, 15(3), 331-377.

ANNEX A ABBREVIATIONS

BILWISS	Bildungswissenschaftliches Wissen und der Erwerb professioneller Kompetenz in der Lehramtsausbildung
CK	Content Knowledge
CME	Classroom Management Expertise
COACTIV	Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz
COACTIV-R	Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz – Referendariat
DIDAKTUM	Didaktisches Wissen und berufliche Motivation von Lehramtsstudierenden
EMW	Entwicklung von berufsspezifischer Motivation und pädagogischem Wissen in der Lehrerausbildung
ETS	Educational Testing Service
GPK	General Pedagogical Knowledge
INTASC	Interstate New Teacher Assessment and Support Consortium
IRT	Item-Response-Theory
KADDS	Knowledge of Attention Deficit Disorder Scale
LEK	Längsschnittliche Erhebung pädagogischer Kompetenzen von Lehramtsstudierenden
MPCK	Mathematical Pedagogical Content Knowledge
MT21	Mathematics Teaching in the 21st Century
OECD	Organisation for Economic Co-operation and Development
PACT	Performance Assessment for California Teachers
PCK	Pedagogical Content Knowledge

PISA	Programme for International Students Assessment
PLANVOLL	Planungskompetenz von Lehrerinnen und Lehrern
PLT	Principles of Learning and Teaching
PPK	Pedagogical/Psychological Knowledge
PTICK	Pedagogical Technology Integration Content Knowledge
SEM	Structural Equation Modelling
SKILL	Studie zur Kompetenzentwicklung in der Lehrerinnen- und Lehrerausbildung für Berufsschulen
TEDS-FU	Follow Up
TEDS-M	Teacher Education and Development Study – Learning to Teach Mathematics
TIMSS	Third International Mathematics and Science Study / <i>Trends in International Mathematics and Science Study</i>
TPACK	technological pedagogical and content knowledge
TPCK	technological pedagogical content knowledge
TSC	thinking in science classroom

ANNEX B: TERM SEARCH RESULTS

No	Title	Author	Publication Date	ERIC #	Empirical	Comment	Select
8	Voices from the trenches: An exploration of teachers' management knowledge.	Garrahy, D. A., Cothran, D. J., & Kulinna, P. H.	2005	EJ725273	yes	classroom management	yes
44	Key pedagogical principles and their major obstacles as perceived by comprehensive school teachers.	Atjonen, P., Korkeakoski, E., & Mehtäläinen, J.	2011		yes	Pedagogical key principles, obstacles, and success	yes
46	Motivations for choosing teaching as a career: effects on general pedagogical knowledge during initial teacher education	König, J., & Rothland, M.	2012		yes	General pedagogical knowledge for teaching (TEDS-M instrument, short version)	yes
54	Teachers' pedagogical knowledge in the instruction of higher order thinking.	Zohar, A.	2002		yes	very explorative	yes
56	The Nature and Development of Teachers' Metastrategic Knowledge in the Context of Teaching Higher Order Thinking.	Zohar, A.	2006		yes	very explorative	yes
64	General pedagogical knowledge of future middle school teachers: On the complex ecology of teacher education in the United States, Germany, and Taiwan.	König, J., Blömeke, S., Paine, L., Schmidt, W. H., & Hsieh, F. J.	2011		yes	General pedagogical knowledge for teaching (TEDS-M instrument)	yes
1	Critical elements for the science teacher to adopt a student-centered approach: The case of a teacher in transition.	Gunel, M.	2008	EJ811892	yes	teacher's perceptions, epistemology, and understandings of student-oriented learning environments - very close to PCK (Science)	no
2	Teaching as a Site of Professorial Work. The Knowledge Base of a Foreign Language Professor from the Perspective of Global Education.	Savukova, G.	2002	EJ652028		rather PCK (Foreign Language)	no
3	Family child care learning environments: caregiver knowledge and practices related to early literacy and mathematics.	Phillips, B. M., & Morse, E. E.	2011	EJ933387		teachers are not target group	no
4	On Pedagogy	Gergely, G., Egyed, K., & Király, I.	2007	EJ849096	yes	teachers are not target group	no
5	Trainee teachers' perception of their knowledge about expert teaching.	Jegade, O., Taplin, M., & Chan, S. L.	2000	EJ614228	yes	maths, science knowledge	no
6	The Numeracy Research and Development Initiative Projects.	Vincent, J.	2004	EJ793908	yes	CK, PCK (maths)	no
7	Investigating experienced ESL teachers' pedagogical knowledge.	Gatbonton, E.	1999	EJ583929	yes	English as a second language => PCK	no
9	Someone else's game: Constructing the English teaching professional in New Zealand.	Locke, T.	2004	EJ731199	no	professional knowledge of English teachers (=> PCK)	no

No	Title	Author	Publication Date	ERIC #	Empirical	Comment	Select
10	Pedagogy in a Public Space: children and adults learning together at Tate Modern.	Ross, M., Hancock, R., & Bagnall, K.	2004	EJ738539		teachers are not target group	no
11	A career in teaching: decisions of the heart rather than the head.	O'Sullivan, M., MacPhail, A., & Tannehill, D.	2009	EJ864905	yes	motivational study (The focus of this research was to understand why young people in Ireland are attracted to teaching physical education and what factors contribute to their decision to enter a teacher education program)	no
12	Identifying mentoring practices for developing effective primary science teaching.	Hudson, P.	2005	EJ721976	yes	rather PCK (primary science education)	no
13	Mathematics primary teachers training at the University of Granada. An adaptation to the EHEA.	Ruiz, F., Molina, M., Lupiáñez, J. L., Segovia, L., & Flores, P.	2009	EJ836624	yes	rather PCK (Maths)	no
14	The Education and Preparation Experiences Principals Seek in Hiring Teachers for Border Middle-Grade Schools.	Desiderio, M. F.	1998	EJ575997	yes	unclear	no
15	Education for sustainable development: Liberation or indoctrination? An assessment of faculty members' attitudes and classroom practices.	Qablan, A. M., Al-Ruz, J. A., Khasawneh, S., & Al-Omari, A.	2009	EJ884405	yes	focus is on attitudes, less on knowledge	no
16	Use of research-based instructional strategies in introductory physics: Where do faculty leave the innovation-decision process?	Henderson, C., Dancy, M., & Niewiadomska-Bugaj, M.	2012	EJ977809	yes	rather pck (physics)	no
17	Becoming a better teacher: A case of changing the pedagogical knowledge and beliefs of law professors.	Hativa, N.	2000	EJ620225	yes	Law Professors' PK, law students' rating; School teachers are not target group	no
18	Investigating the acquisition of pedagogical knowledge: Interviews with a beginning teacher of the gifted.	Joffe, W. S.	2001	EJ632568	yes	biographical	no
19	"I Wash My Face with Dirty Water" Narratives of Disability and Pedagogy.	Gabel, S. L.	2001	EJ620697	yes	biographical	no
20	Application of computer games in the field of education.	Jayakanthan, R.	2002	EJ652458	no (seemingly)	computer games	no
21	The dilemma of dispositions.	Phelps, P. H.	2006	EJ745167	yes	biographical	no
22	Pre-Service Teachers' Perceptions of a Short-Term International Experience Programme.	Barkhuizen, G., & Feryok, A.	2006	EJ729017	yes	"investigated pre-service English second language teachers' perceptions " = PCK	no
23	The development of teaching cases for instrumental music methods courses.	Conway, C. M.	1999	EJ618889	yes	rather pck (music)	no
24	Zwischen Gesetz und Fall. Mutmaßungen über Typologien als pädagogische Wissensform. (Between General Law and the Individual Case. Conjectures Concerning Typologies as a Form of Pedagogical Knowledge).	Herzog, W.	2003	EJ676730	no	not in English	no

No	Title	Author	Publication Date	ERIC #	Empirical	Comment	Select
25	Profiling primary school teachers in relation to art teaching.	Pavlou, V.	2004	EJ686003	yes	rather pck (art)	no
26	Examining instructional practices of elementary science teachers for mathematics and literacy integration.	Douville, P., Pugalee, D. K., & Wallace, J. D.	2003	EJ957605	yes	across subjects, PCK	no
27	Meeting the Challenge of Diversity: Professional Development for Teacher Educators.	O'Hara, S., & Pritchard, R. H.	2008	EJ810649	yes	partly PCK (English); self-report, no testing	no
28	Selection of printed curriculum materials in physical education: Recontextualising pedagogical knowledge.	Devís-Devís, J., Molina-Alventosa, J., Peiró-Velert, C., & Kirk, D.	2011	EJ933223	yes	rather pck (physical education), curriculum knowledge	no
29	Implementing the professional development standards: an innovative MS degree for high school chemistry teachers.	Bretz, S. L.	2002	EJ656461	early evaluation results - unclear	rather pck (chemistry)	no
30	Preserving Music Teaching Field Experiences Utilising an Urban Minority After School Program.	Neill, S. L.	2004	EJ852084	yes	rather pck (music)	no
31	Facilitating dynamic pedagogical decision making: PEPE and GTE.	Wasson, B.	1998	EJ573903	no	conceptual paper on PK (also somehow including other knowledge areas, but the author is not specific).	no
32	Adaptivity and autonomy development in a learning personalisation process.	Verpoorten, D.	2009	EJ871331	no (seemingly)		no
33	Learning to be a science teacher: Reflections and lessons from video-based instruction.	Tan, A. L., Tan, S. C., & Wettasinghe, M.	2011	EJ936634	yes	rather PCK	no
34	PAIR-up	Jorn, L., Whiteside, A., & Duin, A.	2009	EJ833671	no		no
37	From content to practice: Sharing educational practice in edu-sharing.	Klebl, M., Krämer, B. J., & Zobel, A.	2010	EJ901405	no (seemingly)	e-learning	no
36	Competencies of Traditionally Prepared and Lateral Entry Teachers: Implications for School Administrators.	Brown, M. B., Bolen, L. M., Lassiter, C. L., & Burke, M. M.	2006	EJ843070	yes	school principals' perceptions of first-year teachers' competencies / general pedagogy; rating scales to assess future teachers	no
37	Teaching and teacher education: A glimpse into the future.	Kremer-Hayon, L.	1998	EJ578684	no		no
38	Complex Text in ESL Grammar Textbooks: Barriers or Gateways?.	Lesikin, J.	2000	EJ626519		ESL, textbooks, grammar	no
39	How to Change University Faculty Members' Attitudes and Behavior in the Context of Education for Sustainable Development.	Qablan, A., & Al-Qaderi, S.	2009	EJ870317	yes	focus is on attitudes, less on knowledge	no
40	A confirmatory factor analytic approach on perceptions of knowledge and skills in teaching (PKST).	Choy, D., Lim, K.M., Chong, S., & Wong, A.F.	2012		yes	structure of what is defined as GPK; no test, but self-report scales	no
41	A science teacher's reflections and knowledge growth about STS instruction after actual implementation.	Tsai, C. C.	2002		yes	rather PCK (Science)	no

No	Title	Author	Publication Date	ERIC #	Empirical	Comment	Select
42	Beginning teachers' perceptions of their levels of pedagogical knowledge and skills: did they change since their graduation from initial teacher preparation?.	Choy, D., Chong, S., Wong, A. F., & Wong, I. Y. F.	2011		yes	structure of what is defined as GPK; no test, but self-report-scales	no
43	Chapter 7: Assessing pedagogical knowledge.	Metzler et al.	2000		yes	PCK	no
45	Leveraging the affordances of Youtube: The role of pedagogical knowledge and mental models of technology functions for lesson planning with technology.	Krauskopf, K., Zahn, C., & Hesse, F. W.	2012			TPCK	no
47	Presence in teaching.	Rodgers, C. R., & Raider-Roth, M. B.	2006		no		no
48	Preservice teachers' connections of pedagogical knowledge to mentoring at-risk adolescents: Benefits and challenges.	Garza, R. & Ovando, M.	2012		yes	not clearly related to PK	no
49	Representation in teaching: Inferences from research of expert and novice teachers.	Hogan, T., Rabinowitz, M., & Craven III, J. A.	2003		no	GPK (but also partly PCK, CK), Expert-Novice-Comparison	no
50	Researching design practices and design cognition: contexts, experiences and pedagogical knowledge-in-pieces.	Kali, Y., Goodyear, P., & Markauskaite, L.	2011		no		no
51	Science teachers' conceptual growth within Vygotsky's zone of proximal development.	Jones, M. G., Rua, M. J., & Carter, G.	1998		yes	rather PCK (Science)	no
52	Science teaching as knowledgability: A case study of knowing and learning during coteaching.	Roth, W. M.	1998		yes	rather PCK (Science)	no
53	Teacher development: Roles of domain expertise and pedagogical knowledge.	Shulman, L. S.	2000		no (seemingly)		no
55	The construction process of pedagogical knowledge among nursing professors.	Backes, V. M. S., Moyá, J. L. M., & Prado, M. L. D.	2011		yes	PCK	no
57	Understanding teachers' levels of use of learnscapes.	Skamp, K.	2009		yes	unclear, rather PCK	no
58	Schooling and governance: Pedagogical knowledge and bureaucratic expertise in the genesis of the Argentine educational system.	Southwell, M.	2013		no (seemingly)		no
59	Teacher authority, tacit knowledge, and the training of teachers.	Elliott, J. G., & Stemler, S. E.	2008		yes	related to mathematical knowledge primarily	no
60	Models for statistical pedagogical knowledge.	Godino, J. D., Ortiz, J. J., Roa, R., & Wilhelmi, M. R.	2011			rather pck (mathematics)	no
61	Does teacher professional development affect content and pedagogical knowledge: How much and for how long?	Goldschmidt, P., & Phelps, G.	2010			rather pck (English language)	no
62	Preparing teachers through case analyses.	Groth, R. E., & Xu, S.	2011			rather PCK (mathematics)	no

No	Title	Author	Publication Date	ERIC #	Empirical	Comment	Select
63	Pedagogical knowledge representation through concept mapping as a study and collaboration tool in teacher education.	Koc, M.	2012		yes	not directly related to PK	no
65	Teachers' pedagogical competence as a prerequisite for entering the profession.	Liakopoulou, M.	2011		no	conceptual paper	no
66	Facilitating problem-based learning in teacher education: getting the challenge right.	Pourshafie, T., & Murray-Harvey, R.	2013		yes	problem based learning, more related to that approach than to PK	no
67	Teacher knowledge and teaching in Panama and Costa Rica: A comparative study in Primary and Secondary education.	Sorto, M. A., Marshall, J. H., Luschei, T. F., & Martin, C.	2009		yes	related to mathematical knowledge primarily	no
68	Complex Course Generation Adapted to Pedagogical Scenarios and its Evaluation.	Ullrich, C., & Melis, E.	2010		yes	course generation, more related to that approach than to PK	no
69	Professional learning communities: Teachers, knowledge, and knowing.	Wood, D. R.	2007		yes	focus on learning communities, more related to that approach than to PK	no